SANDIA REPORT

SAND2011-9257
Unlimited Release
Printed December 2011

Nuclear Energy Advanced Modeling and Simulation Waste Integrated Performance and Safety Codes (NEAMS Waste IPSC):

Subcontinuum-scale Verification and Validation Strategy

Peter A. Schultz

Prepared by
Sandia National Laboratories
Albuquerque, New Mexico 87185 and Livermore, California 94550

Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.

Sandia National Laboratories

Issued by Sandia National Laboratories, operated for the United States Department of Energy by Sandia Corporation.

Printed in the United States of America. This report has been reproduced directly from the best available copy.

Available to DOE and DOE contractors from
U.S. Department of Energy
Office of Scientific and Technical Information
P.O. Box 62
Oak Ridge, TN 37831

Telephone: (865) 576-8401
Facsimile: (865) 576-5728
E-Mail: reports@adonis.osti.gov
Online ordering: http://www.osti.gov/bridge

Available to the public from
U.S. Department of Commerce
National Technical Information Service
5285 Port Royal Rd.
Springfield, VA 22161

Telephone: (800) 553-6847
Facsimile: (703) 605-6900
E-Mail: orders@ntis.fedworld.gov
Online order: http://www.ntis.gov/help/ordermethods.asp?loc=7-4-0#online

SAND2011-9257
Unlimited Release
Printed December 2011

Nuclear Energy Advanced Modeling and Simulation Waste Integrated Performance and Safety Codes (NEAMS Waste IPSC): Subcontinuum-scale Verification and Validation Strategy

Peter A. Schultz

Sandia National Laboratories
P.O. Box 5800
Albuquerque, New Mexico 87185

Abstract

The objective of the U.S. Department of Energy Office of Nuclear Energy Advanced Modeling and Simulation Waste Integrated Performance and Safety Codes (NEAMS Waste IPSC) is to provide an integrated suite of computational modeling and simulation (M&S) capabilities to quantitatively assess the long-term performance of waste forms in the engineered and geologic environments of a radioactive-waste storage facility or disposal repository. Achieving the objective of modeling the performance of a disposal scenario requires describing processes involved in waste form degradation and radionuclide release at the subcontinuum scale, beginning with mechanistic descriptions of chemical reactions and chemical kinetics at the atomic scale, and upscaling into effective, validated constitutive models for input to high-fidelity continuum scale codes for coupled multiphysics simulations of release and transport. Verification and validation (V&V) is required throughout the system to establish evidence-based metrics for the level of confidence in M&S codes and capabilities, including at the subcontiunuum scale and the constitutive models they inform or generate. This Report outlines the nature of the V&V challenge at the subcontinuum scale, an approach to incorporate V&V concepts into subcontinuum scale modeling and simulation (M&S), and a plan to incrementally incorporate effective V&V into subcontinuum scale M&S destined for use in the NEAMS Waste IPSC work flow to meet requirements of quantitative confidence in the constitutive models informed by subcontinuum scale phenomena.

Acknowledgments

The author gratefully acknowledges numerous insights and contributions developed from conversations and contributions from many individuals in the development of this document: Carter Edwards, who defined the standard for the NEAMS Waste IPSC; Julie Bouchard, whose initial efforts in adapting existing concepts in V&V to subcontinuum inform the current document and who originally was intended have a larger role for a much more ambitious milestone; Xin Sun and Fei Gao at Pacific Northwest National Laboratory, with whom this was originally intended to be a joint milestone document with the Fundamental Methods and Models Program Element of the NEAMS project; and others too numerous to name who have and continue to engage in the ongoing dialogue of what V&V means for subcontinuum scale simulations.

Foreword

This document was originally intended to be much more ambitious in its scope and more inclusive in its contributions from a wider team across NEAMS, but challenging and changing programmatic circumstances have dictated a smaller effort and significantly narrowed scope. The bulk of future subcontinuum scale work on waste forms was transitioned out of the NEAMS Waste IPSC and funding for the joint effort in the Fundamental Methods and Models (FMM) program element was substantially delayed. The plan to develop a V&V plan for subcontinuum scale work jointly between the Waste IPSC and FMM was first postponed and eventually become untenable. With programmatic drivers for the document diminished, good arguments were made that the effort should be either deferred until more favorable circumstances emerged, or simply cancelled. In the end, it was concluded that the conversation of V&V issues in subcontinuum scale scientific studies contributing to an engineering program such as NEAMS needed at least to begin, if only conceptually and generically rather than comprehensively. The current goal is to establish the principle of the importance of V&V and UQ to subcontinuum scale and to start down the path to instill a culture where active consideration of V&V and UQ are an expected aspect of routine scientific work that contributes to NEAMS. This document reflects this much reduced charter.

The original team was to include among its principals Julie Bouchard at Sandia and Xin Sun, Fei Gao, and Ram Devanathan at Pacific Northwest National Laboratory. The conviction that V&V and UQ are important to subcontinuum M&S, and not just within NEAMS, was strongly believed across this team and all brought perspective and contributions that were valuable for defining a useful V&V guidelines for subcontinuum modeling. The failure of the author to substantively include them in crafting the final product reflects only on scheduling pressures and resource limitations. Many of their ideas and insights were incorporated into this document; any inadequacies in the presentation are solely the fault of the author.

So the intent in this is to begin the conversation about strategies to incorporate V&V and UQ principles into subcontinuum scale simulations, and the hope is that subsequent efforts within NEAMS (and without) will expand this conversation into substantive and useful V&V guidelines that will fully incorporate subcontinuum scale science into an evidence-based system of M&S to support risk-informed decision making to advance nuclear energy and science-based engineering.

Let the dialogue begin.

- *Peter A. Schultz*

PAGE INTENTIONALLY BLANK

Contents

Figures

Tables

Nomenclature

CM	configuration management
DFT	density functional theory
DOE	Department of Energy
EBS	engineered barrier system
ECT	Enabling Computational Technologies
EVIM	evidence information management
FEP	feature, event, and process
FMM	Fundamental Methods and Models
IPSC	Integrated Performance and Safety Codes
KMC	kinetic Mote Carlo
MD	molecular dynamics
M&S	modeling and simulation
NEAMS	Nuclear Energy Advanced Modeling and Simulation
NRC	Nuclear Regulatory Commission
PA	performance assessment
PCMM	Predictive Capability Maturity Model
PDE	partial differential equation
PIRT	Phenomena Identification Ranking Table
PNNL	Pacific Northwest National Laboratories
PP	pseudopotential
QC	quantum chemistry
SA	sensitivity analysis
Sandia	Sandia National Laboratories
SQE	software quality engineering
THCM	thermal-hydrological-chemical-mechanical
THCMBR	thermal-hydrological-chemical-mechanical-biological-radiological
UQ	uncertainty quantification
V&V	verification and validation
VVUQ	verification and validation and uncertainty quantification
WF	waste form
WP	waste package

1 Introduction

The objective of the U.S. Department of Energy (DOE) Office of Nuclear Energy Advanced Modeling and Simulation Waste Integrated Performance and Safety Codes (NEAMS Waste IPSC) program element is to provide an integrated suite of computational modeling and simulation (M&S) capabilities to assess quantitatively the long-term performance of waste forms in the engineered and geologic environments of a radioactive-waste storage facility or disposal repository [NEAMSWaste2009]. The goal is to simulate, with quantitative confidence, the long-term release rate into the environment of radionuclides immobilized within a waste form emplaced in a disposal system, beginning with the processes that govern release of radionuclides from a degrading waste form, subsequent reactive transport through engineered barriers, and ultimately into the geosphere. To enable predictive simulation-based assessment, in support of risk informed decisions concerning sequestering and disposing of nuclear waste, requires establishing a defensible level of confidence in the results of simulations.

To meet its responsibility to develop M&S capabilities with assessed levels of quantitative confidence, the Waste IPSC has developed a Verification and Validation (V&V) Plan [NEAMSWaste2011]. This plan defines practices needed to integrate V&V and uncertainty into Waste IPSC activities, enable assessment of acquired and developed M&S capabilities, and maintain and communicate supporting evidence.

To achieve its goals, the Waste IPSC spans three levels of model fidelity: (1) process models developed from mechanistic sub-continuum scale (atomistic) chemical processes, upscaled into constitutive models for use in (2) continuum scale high-fidelity coupled thermal-hydrological-chemical-mechanical (THCM) multi-physics simulations, and further abstracted into surrogate models for use by (3) performance assessment (PA) codes. Every level in this hierarchy must establish a level of quantitative confidence in model predictions, as must the upscaling processes bridging between the scales.

The Waste IPSC V&V plan emphasized applications to high-fidelity continuum scale simulations, the larger focus of Waste IPSC development activities. The scientific codes and simulations comprising the subcontinuum scale are diverse, dynamic, distributed, and fundamentally distinct in character and use from the high-fidelity continuum scale codes. This Report extends the Waste IPSC V&V Plan to address activities at the subcontinuum scale, defining responsibilities and outlining guidelines and strategies for V&V for modeling of atomistic and mesoscale processes.

Subcontinuum activities identify and characterize the crucial rate-determining subcontinuum phenomena, and then assemble and aggregate those phenomena—via upscaling—into predictive mechanistic-based continuum scale constitutive models. Principles of VVUQ for subcontinuum scale activities are usually poorly defined and haphazardly applied, and differ from continuum scale M&S., Subcontinuum-based models must establish credible regimens for V&V to be incorporated into an evidence-based process in the Waste IPSC workflow.

Upscaling itself is not a subject of this Report. Upscaling is phenomena-specific and subject of active research, and the approaches for VVUQ through upscaling are unsolved. This Report articulates VVUQ principles for intra-scale subcontinuum modeling, to provide well-characterized subcontinuum components, appropriately VVUQ-assessed, for use as input into upscaleding.

1.1 Synopsis of a subcontinuum V&V strategy

In distilled form, a viable V&V plan for subcontinuum scale investigations must include consideration of the following components:

1. Establish and document line-of-sight.
   What are the *requirements* and what path do they evolve from?
   What are the *output quantities* of the activity, how are they to be used, what is the level of *required accuracy and uncertainties*?
   What are the needed *input quantities and models* to the activity, where are they obtained, do they satisfy the requirements of the activity or can they be refined?

2. Traceability, reproducibility, and assessability.
   What was done, and how was it done? Any numerical quantities or models contributing to a NEAMS data flow must have their provenance tracked, sufficiently documented for the results to be reproduced, with enough detail that its quantitative pedigree can be assessed.

3. Verification of codes, methods, and models, validation of methods.
   How were the results generated and how credible are they? This includes specification of the codes and models used in the analyses, description of methods to verify the proper functioning of the codes and appropriateness of the models used in the analyses. How will the results be validated? A roadmap for by which the results will be validated in a meaningful way must be described, given that there is often very limited appropriate data available for validation.

4. Uncertainty quantification for output quantities
   What is the level of quantitative confidence in the results? The requires identification of all sources of uncertainties, both numeric and in model form (physical approximations), description of methods to assess and refine uncertainties commensurate with the risks and needs of the domain application, evidence that the numerical uncertainties of the simulations are smaller than the model form uncertainties stemming from the physical approximations.

5. Implementation plan
   How will a regimen of quantitative confidence be implemented in the activities? Outline the realistic and practical evolution of V&V practices within the activity targeted to achieve the levels of rigor commensurate with the requirements of the activity. A roadmap of continuous improvement with tangible, incremental progress from the initial scoping studies through potential model implementation.

Development and implementation of V&V plans encompassing these components is expected from every subcontinuum scale activity. This V&V strategy document provides the guidelines for such a plan, with the intent to facilitate development of realistic, realizable V&V plans for subcontinuum scale efforts, a regimen that add value to the activities rather than simply adding compliance requirements.

1.2 Scope and purpose

The scope of this document is to articulate the principles for V&V and UQ practices to be developed within individual subcontinuum scale application domains. This Report descends from the Waste IPSC V&V Plan [NEAMSWaste2011], extending that framework of broad requirements on M&S activities in NEAMS Waste IPSC into a strategy for implementing meaningful VVUQ at the subcontinuum scale. The scope excludes issues related to upscaling, except as anticipated needs for upscaling impose requirements for quantitative rigor upon the intra-scale M&S.

Adequate VVUQ protocols will differ between codes, and between different applications using the same code. Subcontinuum processes important to waste degradation and radiological release and transport are not fully known, well-defined requirements have yet to be propagated down from continuum scale needs, and the full array of possible methods and codes that might eventually be required is not yet enumerated. The V&V strategy here does not represent a VVUQ plan for any specific code or application, but rather describes expectations and establish guidelines for developing VVUQ plans for subcontinuum scale activities. It outlines a strategy for implementing these guidelines into standard subcontinuum practices within the NEAMS Waste IPSC work flow.

Section 2 describes background of the NEAMS Waste IPSC and the role of subcontinuum scale phenomena within a system model of the Waste IPSC. Levels in the hierarchy are connected through upscaled models. A survey of subcontinuum codes and applications domains is presented, along with the special challenges associated with the upscaling to continuum scale models.

Section 3 articulates general principles and specific practices for V&V and UQ, geared toward subcontinuum scale activities.

Section 4 maps expected practices to a given level of rigor demanded of an activity, using line-of-sight to requirements, and intended use within that line-of-sight, to define a level of rigor and to formulate criteria for the practices needed to achieve that level of rigor.

Section 5 describes the path forward.

The purpose of this document is to articulate V&V and UQ principles pertinent to the intended uses of subcontinuum activities in the Waste IPSC, and provide a roadmap useful for developing and implementing V&V and UQ plans that are meaningful and add value at the subcontinuum scale. It is anticipated that this will be a living document, updated and revised with experience developed from adapting V&V and UQ to specific subcontinuum activities.

1.3 Audience, users, and usage

This plan addresses the needs of many different communities: contributors to subcontinuum scale investigations, those who interact with subcontinuum scale effort in the design of upscaling processes and continuum scale efforts, the VVUQ community, and other programmatic sponsors and stakeholders. The needs of these communities are summarized below.

Subcontinuum scale investigations: to establish a culture of exercising and documenting due diligence driven by the need to establish quantitative confidence for use in engineering systems, and provide a *practical* guide to develop domain-specific plans, driven by value-added proposition rather than compliance. This document is intended to be a living document, growing through contributions from V&V Plans developed for individual subcontinuum domains. Subcontinuum scale users span:

- Within the NEAMS Waste IPSC, studies directed toward identifying and characterizing phenomena with IPSC objectives, e.g., investigations into corrosion mechanisms of borosilicate glass directed toward determining release rates of radionuclides. The goal is to establish a credible basis for the upscaled networks of phenomena comprising validated constitutive models used in subcontinuum scale M&S. This document articulates the strategy for defining and implementing appropriate domain-specific V&V plans.

- Within the NEAMS FMM program element: to specifying standards and expectations for new capability development needed to fill gaps in IPSC capabilities, .e.g, *f*-electron methods of quantum chemistry to model actinides.

- External contributors to and downstream users of an operational NEAMS system: To establish tangible standards and protocols for proposing, defining, and delivering results of subcontinuum activities into the NEAMS Waste IPSC system of work/data flow, crafting and implementing V&V plans conforming to the overall V&V strategy.

Upscaling model developers and continuum scale investigations: to document the expectations for contributions from subcontinuum scale investigations to upscaling methods and the continuum scale.

- Within subcontinuum activities: to describe the need to enumerate specific and assessed/assessable definitions of output quantities feeding upscaling efforts, and input quantities upon which a subcontinuum M&S activity depends.

- For continuum scale users with needs for mechanistic-based subcontinuum-informed constitutive models: to emphasize the importance of performing sensitivity analyses for subcontinuum-resolved processes in constitutive models, to quantify requirements and enable prioritization of subcontinuum activities.

Defining development of VVUQ methods: to better communicate the nature of the needs of V&V and UQ in subcontinuum scale activities. The goal is to provide a quantitative technical basis for understanding the unsolved challenges of propagating quantitative confidence through the filtering process of upscaling, and also to convey a better understanding of the nature of subcontinuum scale investigations, especially the differences from continuum scale M&S.

Experimentalists: to communicate the nature of and motivations for needs of subcontinuum scale investigations with NEAMS. Success of M&S activities within an assessed NEAMS system is dependent upon detailed validation from experiment or observation, yet NEAMS itself generally will not generate or commission experiments or field observations. In general, the V&V requirements described here for M&S activities are equally desirable and needed from experiments as well as from simulations:

- Verification – is the experiment conducted correctly, are the boundary conditions characterized properly, is the correct data being captured?
- Validation – is data being captured a faithful representation of the property being measured, how well does the quantity being measured correspond to the behavior or phenomena being reported?
- Uncertainty Quantification – the level of quantitative confidence in the experimental data describing a behavior is needed to assess the level of quantitative confidence in the M&S capabilities used to evaluate a behavior.

The NEAMS system cannot dictate conditions for experimental studies. The goal is two-fold, to inform potential experimental studies of the specific needs of VVUQ-assessed NEAMS-like investigation, and to specify requirements and guidelines for subcontinuum scale M&S for incorporating experimental data from other sources, for the purpose of satisfying V&V needs of M&S activities within NEAMS.

Other customers: The capabilities enabled by subcontinuum acitivities for the purpose of assessing performance of waste forms in disposal systems are the same as those involved in materials design of waste forms for different waste streams. Subcontinuum analyses will be used to inform NE Waste Form Campaign concerning durability of candidate waste forms for selected waste streams, e.g., to determine and quantify rate-limiting steps in long-term corrosion of borosilicate glass.

NEAMS stakeholders and program management: to understand the challenges of V&V and UQ for NEAMS subcontinuum scale activities, communicate commitment and a roadmap for defining and implementing V&V Plans into subcontinuum activities, and provide an informed basis for defining strategic directions and allocating resources.

2 Background

Section 2 describes the constitution of the NEAMS Waste IPSC and the role of subcontinuum scale phenomena within the IPSC. It describes how requirement flow downward from coarser scales and define significant scientific challenges for M&S at the subcontinuum scale, outlines criteria for determining how those challenges are relevant to Waste IPSC goals, and provides the context for the significant and distinct verification, validation, and UQ challenges represented at the subcontinuum scale.

2.1 NEAMS Waste IPSC top-down overview

The goal of the NEAMS Waste IPSC is to develop a system of modeling and simulation capabilities to evaluate the long-term performance of options for disposal of nuclear waste. In its generic form, as shown in Figure 2-1, a disposal option is characterized by radionuclides immobilized in a durable waste form (WF) material, contained inside a waste package (WP) within an engineered barrier system (EBS) designed to protect the contents from the external environment and impede release of any mobilized species, emplaced within a prepared geological setting. As defined by the U.S. Nuclear Regulatory Commission (NRC) in 10 CRF 63.2, the term *waste form* means the radioactive waste materials and any encapsulating or stabilizing matrix. The term *waste package* refers to the waste form and any containers, shielding, packing, and other absorbent materials immediately surrounding an individual container (NRC 2001).

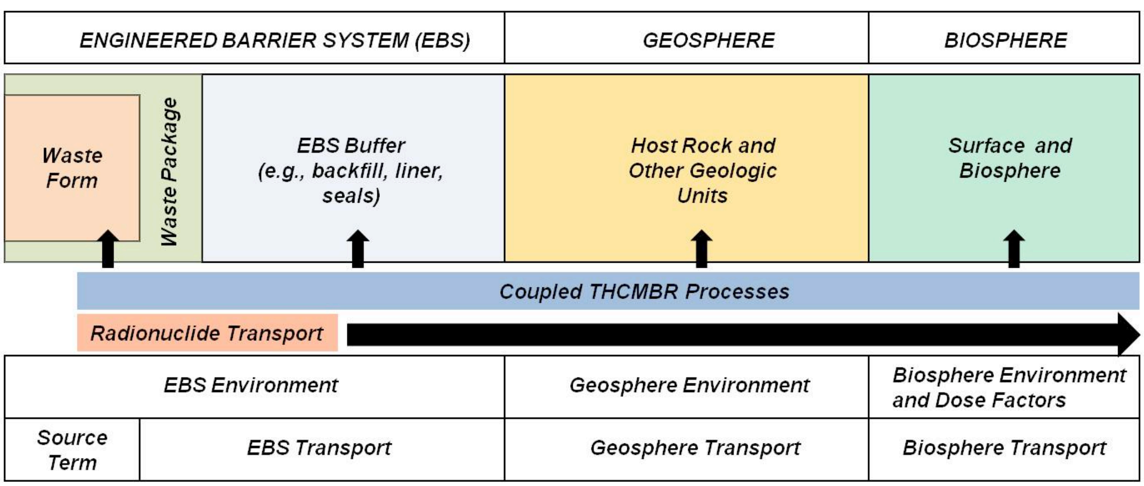

Figure 2-1. Nuclear-waste M&S domain.

Processes leading to ultimate release of radionuclides into the biosphere begin with phenomena that lead to breach of the waste package, expose the waste form to an aqueous environment, causing it to degrade and release radionuclides into the near-field (i.e., within the EBS). The release continues with reactive transport of the radionuclide through the near-field environment and out into the geosphere. The scope of the NEAMS Waste IPSC M&S spans this entire domain, which, in principle, couples multi-physics thermal-hydrological-chemical-mechanical-biological-radiological (THCMBR) processes. The coupling of these processes is crucial in the near-field and declines in importance in the geosphere, as the important behaviors reduce to (relatively) simpler transport processes.

Subcontinuum processes are most important in the near-field region: in phenomena that corrode, crack, and breach the waste package (and EBS), alteration and degradation of the waste form and subsequent release of radionuclides as it comes into contact with an aqueous environment, and modification of the local environment (by the degrading WF and WP) that affects release and transport of radionuclides through the near-field. The NEAMS Waste IPSC will enable simulations for a range of candidate waste form materials—glass, ceramic, metals, or even used nuclear fuel—and engineered barrier systems. Alteration behavior of the WF and breach of the EBS determines the source term for release of radionuclides—containment of the radionuclide within the disposal system. Environmental conditions defined at continuum scales —e.g., thermal and hydrological conditions—determine the boundary conditions for accurate description of the fundamental chemical processes governing materials degradation at the atomic scale. Characterization of materials properties in the WF and WP, identification and quantification of alteration phenomena, and aggregation of these phenomena via upscaling into predictive continuum scale constitutive models is the domain of subcontinuum scale M&S.

Processes that occur at the subcontinuum scale govern the chemistry of WF and WP degradation, but the time scales and length scales amenable to direct mechanistic simulations of atomic processes is picoseconds to perhaps milliseconds, nanometers to microns. The M&S capabilities of the NEAMS Waste IPSC are intended to provide evidence in support of assessments for peak or cumulative dose for a regulator time frame of hundreds or thousands to a million years with an overall length scale of kilometers. These time and length scale are significantly beyond reach of direct subcontinuum scale modeling, the mechanistic atomistic processes must be upscaled into continuum-scale constitutive models that accurately express the collective behavior of the governing chemical processes for larger regions over longer time scales.

The ultimate driver for Waste IPSC activities is supporting licensing applications, the metrics for which are only directly addressable with high-fidelity and PA scale. Subcontinuum scale M&S will not be invoked directly in support of assessing the performance of a disposal system—it is inconceivable that the decision to certify a proposed waste disposal system will hinge on, for example, the outcome of density functional theory calculations of a specific chemical process. However, assessments subcontinuum investigations will materially enable and inform decisions and formulations of models for continuum and PA scale simulations.

The numerical uncertainties from any subcontinuum scale phenomenon may not be explicitly evident in a total uncertainty quantification for an ultimate disposal assessment, but will be implicitly crucial in developing quantitative confidence in the constitutive models used in an overall assessment: inserting the correct chemistry and physics into the constitutive models, and ensuring that the constitutive models accurately reflects that correct physics. Quantitative confidence in the overall PA results is dependent on confidence in the constitutive models used to arrive at those results. Notwithstanding that uncertainties assessed at the subcontinuum scale may not appear in recognizable form in the final uncertainties of the assessment, the crucial role in the design of the models dictates that subcontinuum scale investigations must satisfy same requirement to establish quantitative confidence in predictions as the continuum and PA scale M&S.

Effective verification and validation practices are the mechanism to establish confidence in subcontinuum models, and meaningful uncertainty quantification is needed as a measure of that confidence. Subcontinuum investigations that contribute materially to the NEAMS Waste IPSC must satisfy VVUQ requirements, appropriate to the application domain, and commensurate with the importance of their intended use.

The specific intended uses of subcontinuum are not comprehensively enumerated in advance. Indeed, a significant planned effort and interim goal of the Waste NEAMS IPSC is to identify the phenomena crucial to an overall assessment. Requirements will be propagated downward from a PA scale, where the anticipated regulatory metrics are defined, into finer scale models in a chain of dependencies, but the form this will take is not yet known. The assembly of subcontinuum scale phenomena into collective behavior in the form of effective constitutive models is, in many cases, a scientific research effort; what chemical processes will ultimately prove quantitatively important to collective behavior, i.e. the specific "application" might only determined after significant M&S activities are completed.

The NEAMS Waste IPSC Challenge Problem and its sequence of challenge milestones [NEAMSWaste2010] was promulgated to provide the a tangible application needed to define and implement V&V and demonstrate measurable progress toward development of M&S capabilities. The waste form in this exercise was postulated to be a borosilicate glass. The second challenge milestone involves subcontinuum scale investigations to develop validated constitutive models for the long-term corrosion and dissolution of glass exposd to water. The rate-limiting step in the release of radionuclides embedded in the glass might be: a) chemical dissolution at the glass-water interface, or b) formation of alteration layers and precipitated secondary phases, or c) diffusion and transport of either water into the dissolving surface or d) dissociated species from the surface into the aqueous environment. The relative importance of these phenomena is not known *a priori*. Subcontinuum investigations are used to characterize these processes, and formulate a constitutive model for dissolution rate is constructed (upscaled) based upon these processes. Only then can a validated constitutive model assess the relative importance of these different phenomena at the subcontinuum scale to the overall release rate. Even in this narrowly defined challenge milestone, the exact applications needed to define a conventional V&V plan are only defined after the bulk of subcontinuum activities are complete. Subsequent studies of glass durability might be similar enough to a pioneer study to allow retracing steps through the now-defined path with a predefined V&V plan, but discovery is an inevitable aspect of developing constitutive models.

A V&V strategy for subcontinuum activities must define a process for determining, in a formal manner, what represents the "application", identify the important processes, and outline a procedure for developing effective VVUQ plans for subcontinuum activities to propagate results into intermediate models that are appropriately verified and validated.

2.2 System model for Waste M&S

The NEAMS Waste IPSC, viewed as a system, defines a process to propagate requirements from the PA level, where the overall metrics for performance are specified, down through the continuum scale and ultimately translated into expectations and requirements at the subcontinuum scale. The relationships within this system view are illustrated in the idealized depiction in Figure 2-3**:**

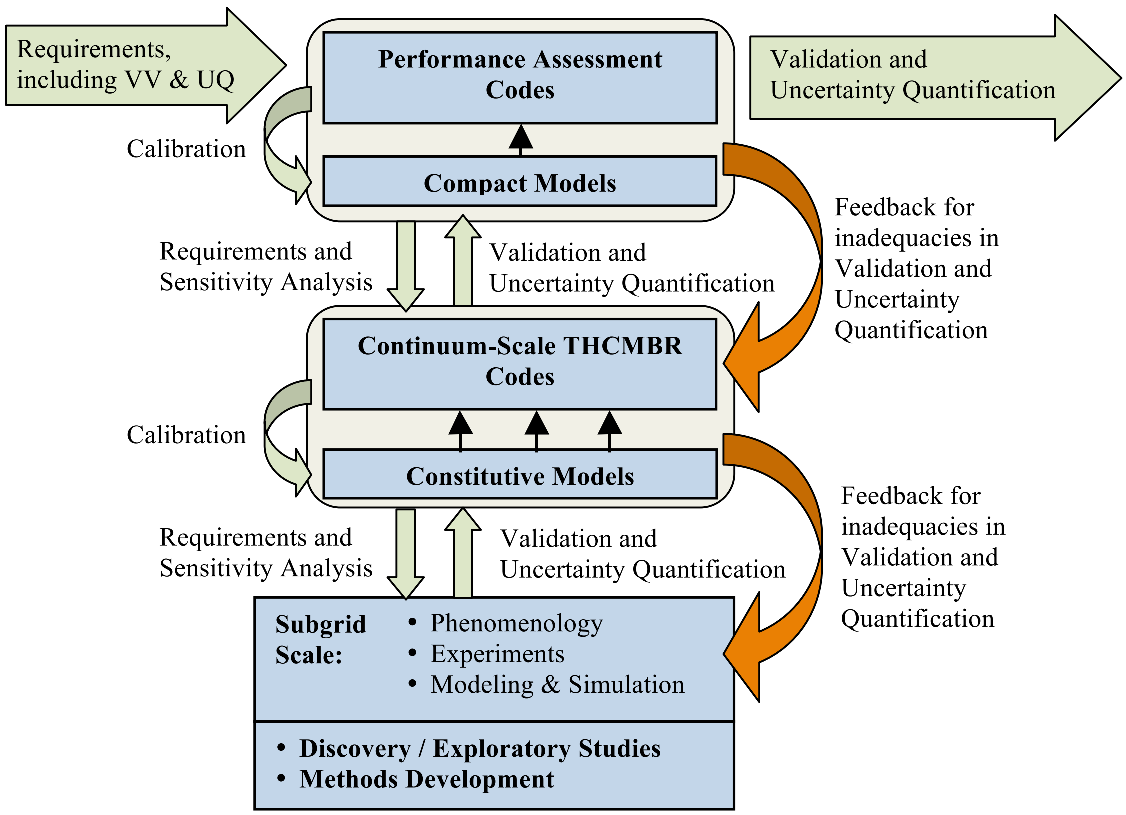

Figure 2-2. System model of Waste modeling and simullation.

The NEAMS Waste IPSC system is composed of three broad layers, corresponding from coarser to finer fidelities: the Performance Assessment scale, the high-fidelity continuum scale, and the subcontinuum phenomena scale.

The feedback between layers shown in Figure 2-3 is crucial to the design of the system. Requirements at one layer determine requirements for the lower layer. The return of validated models and uncertainty quantification from the lower layer may be adequate or inadequate for the needs of the higher layer. Sensitivity analyses passing information downwards identify the most important phenomena, pbenomena that must be more precisely characterized or those which require less focus and precision.

Note that this system could easily be recast as a generic archetype for any M&S-based enterprise that invokes multiscale multi-physics phenomena. The issues pertinent to the Waste IPSC system are the same as many other M&S systems with important phenomena influencing macroscopic behavior being governed by processes at subcontinuum scales.

The layers of the Waste IPSC system model and their inter-relationships are summarized here.

- **Performance Assessment layer.** A suite of PA codes representing fast lower-fidelity, coarser scale, or less-coupled physics is the principal tool for generating evidence for a NEAMS Waste IPSC, designed to run quickly with accurate physics, to assess multiple scenarios and collect statistics for UQ. These codes depend on validated compact models for their accuracy. Inadequacy in validation or inability to meet uncertainty requirements triggers requirements downwards upon the next lower scale, the high-fidelity continuum layer, to reformulate or reparameterize the compact models.

- **High fidelity continuum layer**. The central layer of the Waste IPSC system is a suite of high-performance THCM codes for high-fidelity coupled-multiphysics simulations of continuum scale phenomena. These complex engineering codes are often expressed as solving coupled sets of partial differential equations (PDE). High-fidelity codes typically perform detailed simulations on fine-scale numerical grids incorporating multiple constitutive models describing the response of the various materials to different conditions appropriate to the simulated system. The quality of the constitutive models determines the limits of the accuracy of the simulations, and inadequate validation, or levels of uncertainty that cannot be met through recalibration, signals the need to improve the constitutive models. Reformulating these models extends requirements downwards to subcontinuum scale activities.

- **Subcontinuum layer**. The role of the subcontinuum layer is to evaluate materials properties and mechanistic processes to generate, verify, and validate constitutive models required for use in high-fidelity continuum codes. The anticipation is that empirically-based continuum scale models will be insufficient to extrapolate to reliable, *validated* predictions at PA time scales based available observational time scales, and replacing or augmenting empirical models with more mechanistic models of continuum scale phenomena will enable better extrapolation beyond observation time scales, with greater confidence and lower uncertainties. Upscaling from atomistic scale phenomena to continuum scaling bridges wide time and distance scales. Subcontinuum activities is identify and charactere the crucial subcontinuum phenomena, assemble and aggregate those phenomena through intervening scale internal to the subcontinuum scale ultimately into continuum scale constitutive models. Subcontinuum scale activities respond to requirements propageted downwards from the continuum scale.

- **Interaction between layers**. Interaction between levels in the hierarchy is typically in form of model/parameter passing (sequential) rather than coupled (concurrent) multiscale. Uncertainty quantification passes information upwards in the hierarchy, to propagate uncertainties from finer to coarser scale models. Sensitivity analyses pass information downwards, to identify which phenomena are more important and need to be more precisely characterized and which are less important and need less refinement. A failure to satisfy specified requirements or meet constraints at a given scale, that cannot be mitigated via recalibrating its constitutive models, signals need for improved models, promulgated as requirements downwards to a lower layer of the system. Quantitative weaknesses in these models might be identified through sensitivity analysis (SA). Models at a lower layer generate quantities and models for use in the higher layer, validated and with quantitative uncertainties refined to meet those requirements. Where reparameterization is insufficient, new model forms, to replace or augment empirical models, to incorporate new physics, may be required to reformulate the constitutive models entirely. This feedback—between desired performance at one layer downwards to requirements for the lower layer, and return of validated models from the lower layer with assessed uncertainties—is the process by which requirements are generated down the scales, ultimately into requirements for subcontinuum scale activities.

The distinctions between different layers in this system model are not always be precisely defined, the overlap between PA and continuum scale often being significant (perhaps even using the same codes), and reactive transport simulations of glass corrosion, categorized as a subcontinuum activity, encroaches on continuum scale phenomena, but this hierarchy is, nonetheless, a useful conceptual stratification to understand the roles and relationships of different levels of M&S activities, and the means by which responsibilities are propagated.

This system model is an idealized conceptual structure. The feedback system, in principle, determines all requirements. In practice, this ideal is never fully achieved, requirements and priorities must be specified with incomplete information. This is particularly true for subcontinuum activities. Explicitly connected chains of requirements are not propagated to the lowest levels of the hierarchy in advance. Often, the erquirements are the subject of active research and cannot be known until significant activities have been completed.

Phenomena to be modeled within use of the NEAMS Waste IPSC system are captured in Phenomena Identification and Ranking Tables (PIRT), which are used to identify and prioritize M&S activities. An initial set of PIRTs were developed for the NEAMS Waste IPSC [NEAMSWaste2009]. These requirements evolve in an iterative process, refined as information about the application domain improves. Exercise of the system model is a primary means of refining PIRTs.

The original specification of PIRTs for waste disposal only rarely descended into specific actionable requirements on specific subcontinuum activitie, knowledge is not sufficiently well-developed. A significant goal of the Waste IPSC is to decompose the more (semi-

)empirical descriptions of continuum scale phenomena—representing the current state of knowledge—into first-principles-based subcontinuum scale phenomena. These phenomena can be then aggregated, or upscaled, into more mechanistic and predictive constitutive models that replace or augment empirical models. Requirements on subcontinuum activities are constructing improved *models* rather than results from specific subcontinuum M&S codes and runs. The importance of particular subcontinuum scale phenomena to higher-scale phenomenon is assessed after subcontinuum scale activities have begun and have mostly completed. This complicates the task of defining requirements and establishing well-founded and defensible V&V and UQ plans. The goal of this document is to outline a strategy to develop meaningful V&V plans for subcontinuum scale activities in the face of this intrinsic lack of knowledge.

2.3 Upscaling and models

"Upscaling" is defined as the process of aggregating finer scale processes and phenomena into models for use in simulating coarser scale models. This can occur through statistical averages, reduced order models, or abstraction of physics into simplified form, but all have the net result of a simplified model with fewer degrees of freedom that allow larger scale phenomena be modeled for longer time scales. First-principles-based, mechanistic, validated constitutive models are the output of subcontinuum activities. Requirements propagating down the system model hierarchy are expressed in requirements for models, and a V&V strategy must be expressed in response to those requirements.

Description of V&V for upscaling, and propagation of uncertainties through the filtering processes of upscaling is outside the scope of this Report. However, a formal understanding of upscaling and models is crucial to define useful V&V plans for subcontinuum M&S—validated upscaled models are the product that subcontinuum activities, that will define the requirements for subcontinuum activities.

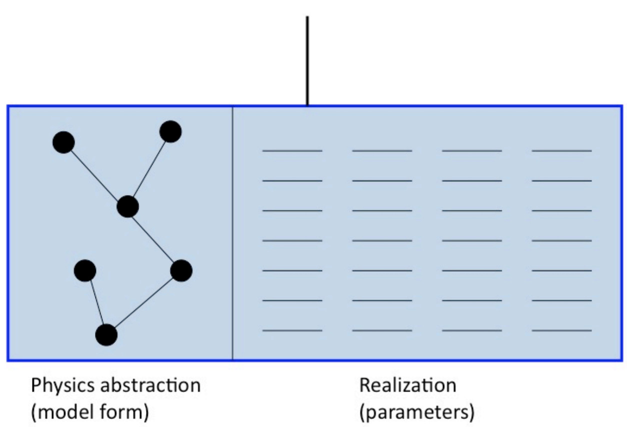

Figure 2-3. Anatomy of an upscaled model: model form and parameters.

A "model" has two major components: a model form, that distills the physics abstraction into mathematical (computable) form, and a realization, the parameters that populate the model to provide quantitative descriptions of materials behavior.

Subcontinuum activities are required for both aspects of this model development: in refining the model form, identifying and assessing the important chemical processes governing a continuum behavior; and in refining the realization of the model, characterizing those processes and populating the parameters of the model with validated quantities. Refining model forms requires identifying and assessing importance of phenomena that are not already known, i.e. not in a PIRT or defined requirement, for if they were known the model would be complete and not need further refinement.

A model ranges between two limits. At one limit, empirical (or phenomenological) models are fit and calibrated to reproduce observed behavior within specific range of material or environmental conditions. At the other limit, subgrid-aware models are explicitly derived from processes at lower scales, striving to reproduce observed behavior through mechanistic interaction of subcontinuum processes. Empirical models are usually interpolative, valid only within the range of conditions spanned by the observational data used to generate the model. Mechanistic models are potentially extrapolative, expressing fundamental principles that are presumably valid outside the range of observations for the coarser scale phenomena.

Typical models represent some intermediate state of knowledge between the purely empirical and the purely mechanistic-generated, a mix of phenomenological description and mechanistic subcontinuum processes in a total semi-empirical model.

Practical semi-empirical models refit and calibrate of the entire model to achieve validation at a given scale. That recalibration will usually extend to the nominally mechanistic-derived aspects of the model. The total model, recalibrated to achieve best validation at the continuum scale, may make use of the mechanistic-inspired elements of the model to overcome other shortcomings (incompleteness) in the model. The parameters in a recalibrated model may take values substantially different from the physical values that correspond to the physical processes they are associated with in the model form. The more complete the model, the better the physical parameters of subcontinuum processes will align with calibrated model parameters. Generally this alignment will be imperfect.

The imperfect correspondence of the numerical requirement does not obviate or diminish the requirement to establish quantitative confidence in the results of subcontinum M&S of those physical phenomena. Those numerical results not only provide numerical parameters, but also support for a level of confidence in the model form. The upscaling process of propagating first-principles-informed information, with uncertainties, from the atomic scale to the continuum may not be perfectly defined. This lack of definition means traditional propagation of requirements into V&V and UQ plans need to be augmented to apply to subcontinuum activities that operate in more ambiguous circumstances.

2.4 Subcontinuum scale survey

The range of activities potentially required within the subcontinuum layer is wide, the potential user community is diverse, and the nature of the tools and codes is similarly broad. A comprehensive V&V strategy is needed to serve these uses and users, enabling viable and meaningful V&V plans to be tailored to specific applications.

2.4.1 Uses and users of subcontinuum codes

Subcontinuum scale information used within NEAMS Waste IPSC system will be obtained from a variety of sources. Subcontinuum activities will be associated with the NEAMS Waste IPSC itself or with the FMM (or other) program element within NEAMS during the course of NEAMS capability development. After the NEAMS Waste IPSC capability is developed and deployed, subcontinuum activities to develop materials-specific models will be commissioned by downstream NEAMS system users for assessing a disposal system. Subcontinuum information needed for model development will be obtained from sources external to NEAMS: active collaborative efforts within NE, such as with the Waste Form Campaign, or other colleagues in basic science research, or passive interactions such as data collection from literature and other records. A viable V&V strategy for subcontinuum scale phenomena invested in developing validated constitutive models must encompass all these potential sources for subcontinuum data.

The emphasis of subcontinuum activities within the NEAMS Waste IPSC is development of models rather than development of subcontinuum codes. The NEAMS Waste IPSC will generally not develop subcontinuum scale codes, except for incidental development in the normal course of subcontinuum investigations. Scientific material science and chemistry communities have vibrant efforts in modeling and simulation, and a vast array of simulation codes are available across subcontinuum scale domains. Any NEAMS related subcontinuum M&S activities will use capabilities previously developed within research communities. Any significant gaps in code capability requirements that cannot be leveraged from the scientific community would be developed in coordination with the FMM program element, but the larger proportion of subcontinuum simulations would use existing codes that NEAMS will not have developed nor will have substantive control over, and may not even have access to source code. One immediate consequence is that software quality engineering (SQE), traditionally an important aid in developing verified simulations codes, will play a more limited role in verification of subcontinuum scale M&S. This shifts the onus of verification more strongly onto the users rather than the developers of capabilities for subcontinuum M&S. A viable V&V strategy for subcontinuum activities must incorporate this reality.

2.4.2 Sources of subcontinuum codes

The NEAMS Waste IPSC will generally not be developing subcontinuum scale codes, nor in any practical sense be dictating which codes will be used. Simulation codes commonly used in subcontinuum scale sciences, and which are within the universe of codes that could by employed to support NEAMS Waste IPSC needs for subcontinuum M&S, range from highly-polished integrated suites of simulations tools with sophisticated user interfaces from commercial vendors, to distributed open-source community code development projects, to individual investigator codes developed on desktop machines with minmal documentation—perhaps even lacking a name.

An acquired code from this universe of codes will have a different degree of "quality"—extent of documentation, test suites, configuration management, literature, etc. Each code, regardless of its provenance and perceived quality, for each application in support of NEAMS Waste IPSC, must be assessed for the purpose it is applied to within NEAMS.

A V&V strategy for subcontinuum M&S activities contributing to NEAMS Waste IPSC, using any of this range of codes, needs to encompass any of these contingencies.

2.4.3 Domains of subcontinuum codes

Subcontinuum phenomena govern degradation of waste forms and waste packages, and contribute to transport and release of radionuclide. Chemical bond-breaking at glass surfaces exposed to water is described by quantum mechanics, transport of water to the pristine glass interface or detached species from the glass into the aqueous environment is governed by diffusive processes described with molecular dynamics. Dissolution of a surface to extract a release rate involves a statistical conspiracy of many bond-breaking and detachments from a surface, perhaps best simulated by kinetic Monte Carlo approaches. Evolution of phases, such as driven by decay of radionuclides in ceramic waste form, might require any of a range of mesoscale tools. Assessment of the importance of different mechanisms to glass corrosion and prediction of long-term rates of dissolution will use reactive transport simulations.

There are different types of simulation methods appropriate to simulate each of these different processes, and for each of these simulations methods, there can be a multitude of codes available and commonly used. The following lists types of simulations and examples of codes associated with each type:

- Quantum chemistry (molecular)
  Gaussian, GAMESS, NWChem, …

- Density functional theory (solid state, plane wave and full-potential):
  VASP, ABINIT, Quantum-Espresso, CRYSTAL, SeqQuest, SIESTA, …

- Molecular dynamics (classical)
  LAMMPS, DLPOLY, MOLDY, GULP, SPASM, … and some home-grown

- Kinetic Monte Carlo
  SPPARKS, ALSOME, … and many home-grown

- Kinetic microcontinuum (KmC) reactive transport
  CrunchFlow, …

- Mesoscale – dislocation dynamics, phase field, Potts model, plasticitiy, grain evolution
  MDDP, ParaDis, … mostly home grown …

This list is not comprehensive, nor meant endorse the listed codes nor disqualify codes not mentioned. The list derives from an informal survey, initially from a NEAMS FMM workshop, of subcontinuum scale tools already in use within broader NEAMS activities. While the phenomena and materials that are the subject of subcontinuum activities will differ between different IPSC project elements, the M&S toolset at the subcontinuum scale used for the different project elements will significantly overlap.

This list is meant to illustrate the wide extent of tools that are used in NEAMS subcontinuum activities, that must be brought within a V&V regime for insertion into a NEAMS Waste IPSC process work flow.

2.4.4 Upscaling within subcontinuum scales

The subcontinuum level will contain multiscale stratification, embedding upscaling processes akin to the upscaling in the Waste IPSC system model (Fig. 2-3), as in Figure 2-4.

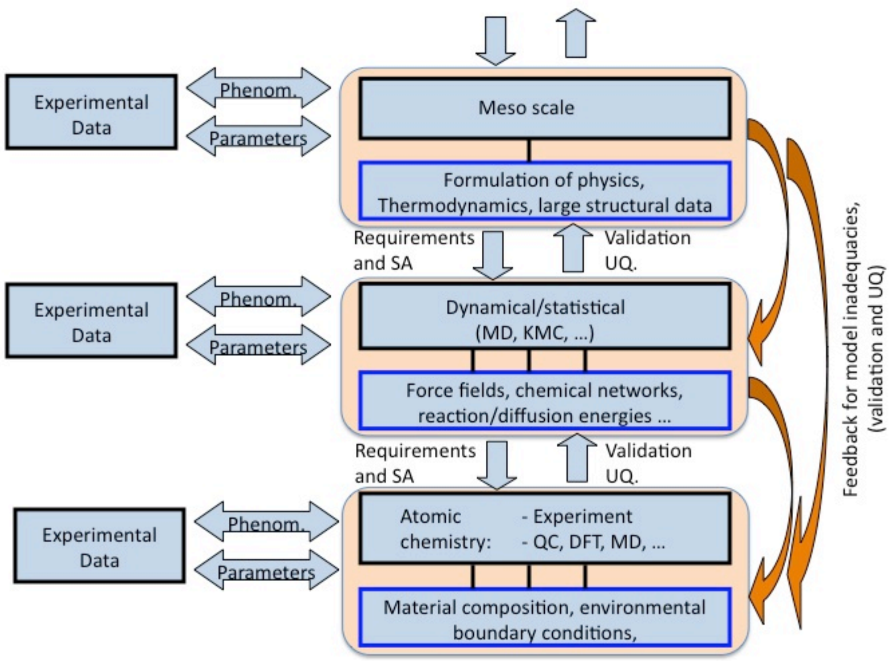

Figure 2-3. Upscaling within subcontinuum scales.

Atomistic, statistical, and mesoscale processes are interconnected in a feedback process of cascading requirements and upscaled models, and a standard pattern for upscaling. Propagating uncertainties through subcontinuum upscaling is an active area of research. This diagram illustrates the information flow, and illustrates the nature of specifying the requirements on individual domains within subcontinuum activities.

3 Generic V&V requirements for subcontinuum

The development of predictive constitutive models that are valid outside the range of field observation will require establishing quantitative confidence in results of subcontinuum scale M&S capabilities used to inform those constitutive models.

Subcontinuum scale activities will be often be undertaken before well-defined requirements are propagated down to the subcontinuum scale, and before constitutive models are fully developed and assessed that allow practitioners to rank the importance of the subcontinuum phenomena being assessed. Subcontinuum M&S activities employ a wide range of codes, from a wide range of sources, by a wide range of communities. Despite the difficulties presented by this heterogeneous enterprise, and the imprecise and dynamic nature of requirements, subcontinuum activities must be and will be subject to verification and validation and quantification of uncertainties appropriate to the nature of calculations, commensurate with their importance to developing validated constitutive models. Thus section describes the concepts necessary to imbue the results of subcontinuum activities with quantitative confidence, adapting conventional concepts to the particular circumstances of an application. This section outlines principles and a strategy for defining V&V requirements for subcontinuum scale activities.

In this section we recap the concept of a level of rigor, an element of the Predictive Capability Maturity Model (PCMM), used as an organizing principle for defining a V&V strategy, and then outline the principle elements of a strategy to define V&V plans for a subcontinuum activity:

1. Establish line-of-sight, to define requirements, identify quantities of interest
2. Establish traceability and reproducibility, enable assessability
3. Establish V&V practices, to define how confidence is developed
4. Establish sources of uncertainties, and assess commensurate with rigor.
5. Establish realistic plan for implementation

3.1 Levels of rigor

The level of confidence in a modeling and simulation capability, including at the subcontinuum scale, is a consequence of the level of rigor to which verification and validation is performed. The Predictive Capability Maturity Model (PCMM) for Computational Modeling and Simulation [Oberkampf2007] was cited in the NEAMS Waste IPSC V&V Plan [NEAMSWaste2011] as an example of a classification scheme for assigning rigor levels to the risks associated with the intended use of particular M&S activities.

Table 3-1. Levels in the PCMM

Rigor Level	Risk Level	Example Usage
0	Low consequence	Scoping studies
1	Moderate consequence	Design support
2	High consequence	Qualification support
3	Highest consequence	Qualification or certification decisions

Higher levels of risk typically imply a need for higher levels of rigor. Higher levels of rigor typically require greater effort and allocation of resources. A PCMM classification is useful as a basis for making decisions concerning resource allocation and prioritization of activities, a system for recalibrating level of effort and minimizing overall risks.

For many PA and continuum scale M&S activities, the required level of rigor is anticipated to be high, as direct output of these simulations might be used directly as evidence to support high-consequence decisions such as licensing. The path to these high-level requirements from subcontinuum scale activities is not as direct, often ill-defined, and the PCMM in Table 3-1 is less useful as a guide for making decisions regarding prioritization of activities, and distinctions of levels of rigor associated with those activities. As stated in the introduction, the ultimate decision to license a waste repository will almost certainly never hinge on the result of any single subcontinuum scale calculation, and the contribution to the level of uncertainty of an overall disposal assessment from the uncertainty in any subcontinuum calculation will be obscured through filtering through multiple upscalings.

For the purpose of being useful for guiding subcontinuum scale activities, a subcontinuum level PCMM needs to be recalibrated to the intermediate scale goal of creating validated constitutive models rather than the indistinct requirements of the overall disposal assessment. The proposed restatement of a subcontinuum-appropriate PCMM is presented in Table 3-2:

Table 3-2. Levels in a subcontinuum-appropriate PCMM

Rigor Level	Risk Level	Example Usage
0	Low consequence	Scoping studies (discovery)
1	Moderate consequence	Design support (development of model form)
2	High consequence	Model qualification (parameterization and validation)
3	Highest consequence	Quantified impact on continuum scale simulation

This subcontinuum scale PCMM is an embedded hierarchy within the NEAMS Waste IPSC system level PCMM. Scoping studies, level of rigor 0, represent science investigations to discover relevant phenomena and assess feasibility of modeling phenomena with particular methods, with relatively few formal requirements. These would generally be outside the scope of NEAMS activities, except as incidental to model development. Model development activities to investigate potentially pertinent phenomena for inclusion to a candidate constitutive model would lie at rigor level 1. The purpose, to screen and assess importance of phenomena, requires some minimum level of confidence in the M&S results used to make those assessments. With the specification of model form and included phenomena, activities to support the parameterization and validation of a model intended for use in continuum scale M&S constitute rigor level 2, and trigger additional V&V requirements. In the event that the physical importance and quantitative impact of a specific subcontinuum phenomenon has a definable and measureable impact on continuum scale phenomena, the level of rigor 3 places the most requirements for V&V activities, and promotes the activity to a ranking (level of rigor) on the overall system PCMM, with all of its attendant requirements for V&V and UQ.

The subordinate subcontinuum PCMM represents a useful classification for making decisions for subcontinuum scale activities and defining a set of appropriately pitched standards for V&V associated with those activities. Verification and validation can only be defined for specified code for a specified use. A V&V strategy for subcontinuum needs to establish a mechanism for defining requirements, i.e., the specified use, in the absence of a fully developed pre-existing set of requirements propagated downwards from the performance assessment scale, The line-of-sight principle is used to delineate a working definition of requirements, as a basis for outlining a V&V plan, in the absence of well-defined requirements explicitly propagated downwards from the PA scale.

3.2 Line of sight

The initial step of a subcontinuum activity, and development of a V&V plan, is to establish the working requirements around which the subcontinuum activity is organized. What is the problem being solved, and what is its path into system requirements, the *line of sight*. What is the constitutive model being developed, what is its role and importance in the system model, and how important is the data being generated to specifying that model? In the absence of a fully-developed roadmap of requirements cascading down from PA scale, the "requirements" must be outlined within the line of sight of the subcontinuum activity. Only after this specification (which will evolve as better information is propagated, perhaps as a result of the subcontinuum activity itself) can an appropriate level of rigor be determined and associated standards of V&V be developed.

A line-of-sight determination must address such issues as …

- *What exactly is the output quantity (or quantities)*?
  For much of scientific research, results are expressed as figures or diagrams or even movies illustrative of physical processes, and can be very insightful. For an assessed engineering work flow, and passing of information between scales, the outputs of simulations need to be cast as quantitative measures, data to be

exchanged between codes and scales, and for which quantitative assessments, for validation and uncertainty quantification, are possible.

- *Into what model does this output quantity go*? *Who will use this quantity*?
  This is the presumptive path retracing requirements upward along a line of sight toward PA scale requirements. Who will use the data, what is the node on the path to the PA scale that will specify requirements on the subcontinuum? This user and node along the line of sight may not yet exist, but the presumptive path must be specified and must be demonstrated to plausibly lie along a path of requirements consistent with the best current state of knowledge of the system. With further development of the systems-level requirements, this path may prove to have less consequence and be deprecated, or may prove more consequential and require additional refinement, but a presumptive path must be specified.

- *What are the necessary input quantities and conditions*?
  Establishing line of sight downwards is as important as upwards. How will meaningful and useful inputs to the simulations be identified and obtained? Are those inputs available, commensurate with the needs of the application, who will provide them? What are the material compositions, environmental conditions, and input quantities to the subcontinuum activities, consistent with the presumptive system level requirements?

- *What associated activities are needed for the model*?
  Frequently, a given subcontinuum activity will address but one aspect of a larger coarser scale model. To repeat an earlier example, release rates from degrading glass will be a function of the overall corrosion rate, that will have contributions from bond-breaking at the glass surface, formation of alteration layers, transport of water and dissolved species to and from the surface, ion exchange, and formation of secondary phases. The overall success of the model for release rate depends on integration of the effects of multiple subcontinuum scale phenomena.

- *What is the state of information for the existing model*?
  Is the existing phenomenology already sufficient? Is there experimental data available to populate the same aspects of the model as the M&S is intended to compute, or how difficult would it be for experimental activities to define this aspect of the model? Is the modeling capability going to be sufficient? These questions form a requirements triage for the subcontinuum activities, illustrated in Figure 3-1. If the existing model is adequate, or the output quantity can be obtained from good experimental data, additional modeling is less important.

These considerations are a predicate for preparing a V&V plan for subcontinuum modeling. They identify working requirements, the role of the activity within the overall system requirements, and enable assignment of a level of rigor to the activity. With the output and input explicitly specified, and a proposed local line-of-sight path of requirements, these specifications then provide a well-defined foundation for upscaling, and more clearly outline the chain of requirements for the overall system.

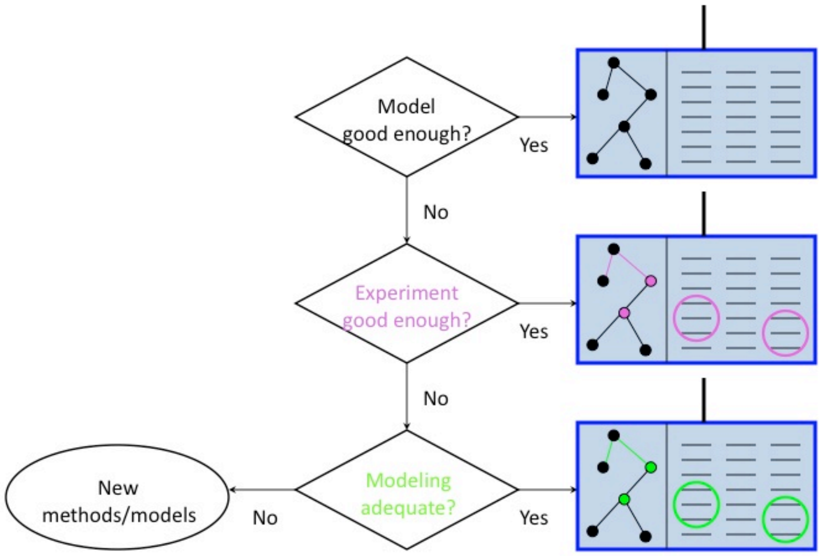

Figure 3-1. Requirements triage for determining flow of requirements.

3.3 Traceability and reproducibility

A minimal requirement of including the results of any subcontinuum activities into the assessed data/process flows in NEAMS is documentation of sufficient knowledge concerning the activity to enable independent assessment of the confidence to be placed in the results. The results must be traceable, with the provenance of the data and associated evidence recorded, with sufficient granularity and detail for the results to be independently reproducible. Any subcontinuum activity in support of NEAMS of must provide documentation to meet this minimal goal, and a V&V plan for that activity must spell out the methods by which this requirement will be satisfied.

This broad principle applies to all subcontinuum data acquisition contributing to NEAMS, including, and, perhaps especially, experimental measurement data acquired from external sources. Validation, essential to NEAMS M&S activities, will be dependent upon this measurement data. Practices for acquiring, assessing, and recording experimental data were described in the NEAMS Waste IPSC V&V Plan [NEAMSWaste2011], and will not be repeated here. In general, the set of practices outlined broadly apply to other (M&S) data obtained from external sources. All subcontinuum scale activities within the NEAMS need to adhere to similar levels of record-keeping requirements.

The goal of much of scientific research at the subcontinuum scale is to *describe* physical phenomena, while the goal of subcontinuum activities within NEAMS is to *predict* behavior. The level of description required to provide sufficient documentation for the latter is much greater than is conventionally provided in published scientific literature with the former. Computational capability and an evolving sense of the meaning of due diligence in science are causing this gap to narrow. A V&V plan for subcontinuum activities must be aware of this distinction, and make provision to satisfy the usually more stringent reporting requirements in an assessed engineering data/process flow.

Different subcontinuum M&S will have varying amounts of information needed to reproduce and enable assessments of calculations. These protocols are rarely carefully documented in public literature, although early forays into establishing reporting standards for DFT simulations for materials science [Mattsson2005] and MD calculations [Kuksin2005, KIM2011] exist to provide guidance into issues that need to be considered in particular applications of those methods. Every subcontinuum activity within NEAMS must outline the crucial issues associated with its M&S methods and specific application, with sufficient granularity and detail to enable an independent assessment.

The type of information that this might entail reporting includes:

- Code and simulations methods: name, version, provenance, acquisition. If available commercially: vendor/version/reliable access, public codes: name/version/download information, home grown codes: describe method for access, within source or in obtaining equivalent. The goal is to enable obtaining the code capability independently so that simulations can be reproduced.

- Model assumptions: the path by which results of the code are mapped onto predictions of phenomena behavior, assumptions and analysis that convert the calculation results into ultimate simulation results.

- The initial conditions: material composition (atomic), structure (e.g., whether a lattice parameter in a DFT materials simulation is the experimental or theoretical value), grain structure, interfaces, etc.

- Computational input parameters and settings: specification of internal grids (e.g., meshes for grid-based codes, real-space grids or reciprocal space sampling grid in a DFT calculation), modeling cell domain (e.g., molecular structure in QC or supercell specification in DFT), interaction or integral cutoffs, basis set specification (e.g. energy cutoff in plane wave DFT, or basis set quality in local orbital DFT).

- Model form inputs: interatomic potentials (for MD), pseudopotentials and functional forms (for DFT), model form and associated parameterization for mesoscale. How is the physics being simulated formulated?

- Tracking of input files, intermediate results, and output files: sufficient to document the choices made in designing the simulations, and connecting those choices to the quantitative output of the simulations.

Subcontinuum M&S activities can require immense quantities of data to set up, and can generate correspondingly immense quantities of data as output. The requirement of reproducibility does not imply the reporting or recording the entirety of this voluminous record. The requirement of reproducibility means documenting the discriminating choices made in the design of the simulations, those choices that can materially affect the results and cannot be unambiguously inferred from the nature of the calculation (not based on the reputation of researcher, or guesses as to likely practice), sufficient to allow a deterministic reconstruction of the essential simulation result. Once again, this criterion will be satisfied differently in different subcontinuum M&S domains. Any subcontinuum activity contributing to NEAMS will need to specify and justify the minimum reproducibility requirements in a V&V plan specific to its domain and application, and document these in any contribution to the data/process work flows.

These requirements can be satisfied in the form of:

- Published literature, in sections describing computational details and methods, or in publicly accessible supplementary material.

- Reports prepared for NEAMS as part of planned activities and milestones

- Synthesized from literature, reports or collaborative activities, as described in the Data Acquisition Activities/Practice section of the Waste IPSC V&V Plan.

- Deposition of data in a NEAMS Evidence Management System.

The purpose of traceability and reproducibility is not V&V and UQ assessment, but to ensure assessability. The first component in outline a V&V strategy, line-of-sight, can be interpreted as the specifying the "why?" of a subcontinuum activity, traceability and reproducibility is the process of documenting the "how?" of generating a result from subcontinuum M&S.

3.4 Verification and validation

Verification and validation (V&V) are practices used to establish quantitative confidence in M&S activities, and this includes the subcontinuum activities invested in developing, refining and qualifying constitutive models for use in continuum M&S. Slightly paraphrased, verification and validation are typically defined as …

- Code verification is the process of the implementation (and execution) of a code that is free of coding errors, is compiled and linked without errors, and executes and solves the mathematical model it is designed to solve, i.e., is a faithful expression of the physical approximations it is designed to implement, with numerical methods that are within the bounds of required precision.

- Model validation is the process of demonstrating that the results obtained from an M&S activity, correctly implemented and executed, can be used for its intended purpose with a specified degree of quantitative confidence.

For the purposes of subcontinuum activities, definition of verification and validation is distilled yet further, to better focus guiding principles for outlining a V&V strategy for developing meaningful and useful V&V plans for specific subcontinuum activities:

- Verification: Is the physics implemented, executed and evaluated correctly, with numerically reliable results?

- Validation: If the physics is implemented and run correctly, do the results give a sufficiently accurate description of reality, i.e. physically reliable results?

Moreover, these principles must be instituted for each component of the subcontinuum activities directed toward a particular output. A single monolithic calculation with a single code rarely constitutes an entire simulation of desired properties and behavior. All elements of an analysis, with multiple calculations at its core, need be adequately assessed to obtain the requisite quantitative confidence in the entire analysis resulting in the desired output.

These distinctions from the conventional descriptions of verification and validation are important for subcontinuum scale activities, as the M&S activities invested in developing the output, improved constitutive models for continuum scale simulations is better and more usefully described as a system of codes, simulations, and analyses, than it is as the single output of a single code or calculation.

This model-centric rather than code-centric view reveals a more expansive set of issues that are subject to verification. Each code or analysis needs to be verified. To establish confidence in the numerical solutions applied to given problem is the realm of solution verification. In addition to these conventional practices, subcontinuum activity needs to explicitly address model verification, the manner in which the results of subcontinuum calculations are processed or analyzed to predict the physical quantity they are intended to simulate. Subcontinuum codes tend to be implemented for general-purpose use, and are not configured to model the exact phenomenon they are applied to. For example, much of the localized chemistry or materials defect phenomena are modeled using DFT codes that are implemented around the assumption of periodic boundary conditions, and the interaction between its periodic replicas needs to be removed or accounted for to accurately describe local chemical processes. Or boundary conditions appropriate to the physical system need to be incorporated into the simulation, for example, when the effects of pH on energy barriers to break bonds at a glass-water interface are desired from a DFT calculation with (once again) periodic boundary conditions, without explicit accommodation of pH within the DFT calculation itself. The model and analysis that connects the calculated values with the predicted values associated with the simulated process need to be verified.

Issues associated with validation take on slightly different connotations as well. The overall constitutive model, the ultimate deliverable of the subcontinuum level activities, often cannot be validated at the subcontinuum scale, it is validated and recalibrated at the coarser (typically continuum) scale. The model form and parameters are used and assessed for suitability of purpose—validated—at the coarser scale. To establish confidence in the proposed model form, subcontinuum scale predictions and observations must be piece-wise validated for subcontinuum processes, building a system of evidence that enables the subcontinuum-scale model components to be adequately assessed.

3.4.1 Code verification

The NEAMS Waste IPSC will generally not develop subcontinuum scale codes, nor will it be anointing or certifying subcontinuum codes for use. Consequently, the onus for code verification with subcontinuum scale activities within Waste IPSC assessments will rest on a user of a code rather than developers of the code. Absent control of the source code, or perhaps even access to the source code, appealing to enabling practices such as software quality engineering (SQE) to facilitate verification assessments is not possible. Assessment of the faithful implementation of the mathematical methods must be accomplished through verification Turing tests: tracking of the codes and executable used, and discriminating assessments that adequately confirm the behavior of the code satisfies conditions of a verified code for the intended uses. Without access to source, verification *failures* often cannot be repaired, but evidence of verification success can be demonstrated. Required code verification practices consist of three broad categories: unambiguous identification of the code and establishing provenance, successful compilation and demonstration of proper functioning and suitability for purpose, and demonstration of adequacy of numerical methods for the application.

3.4.1.1 Code identification and provenance

The first requirement, descending from reproducibility, is to unambiguously identify the code and its provenance, the nature of the platform environment that affect the code compilation or running that possibly affects the results produce by the code. This first category is focused on *identification*, to enable *reproducibility*.

- **Code identification: Code name, version, revision number**
  Sufficient to unambiguously identify code and enable reproduction of simulations.
- **Code provenance: how, where obtained; configuration management**
  Vendor or site where code obtained or provided, or, if source code is recorded in NEAMS evidence management storage, location or tag information, Tangible expression of configuration management, sufficient to enable access to same code and obtaining the same results for the same simulations. Implicit in this practice is that use of home-grown codes must satisfy the same requirement, and must have identification (names/versions) and configuration management sufficient to enable code verification.
- **If platform-dependent: platform, operating system, compilers, libraries used**
  With modern complex codes, building executables can be a significant challenge and the functioning of a code system dependent upon the system configuration. This may be as basic as whether a code was built as MPI-parallel or as a serial application. The operation of a code might depend on the operating system or the selection of libraries, and, in that event, those options must be documented.
- **If jobs are environment-dependent: specification of job environment**
  If results are potentially dependent on the number of processors, or whether coprocessors (such as GPU's) are used, or run as part of a larger code system, those options in running the code must be specified.
- **If numerical method dependent: specify numerical methods and algorithms**
  Nature of meshes, or use of numerical solvers, or other numerical or physical

approximations to the exact solution should be specified. For example, in a DFT calculation, the functional, whether and what type of pseudopotentials are used, or whether the basis set for the solutions is plane waves or local orbitals are among the fundamental approximation that must be identified.

The fundamental requirement is to specify the nature of the calculation(s) sufficiently for them to be reproduced, without requiring querying of the calculation author, perhaps someday unavailable or unable to confirm or corroborate the original simulations.

3.4.1.2 Code performance, numerical model correctness

Having built the code for a specific platform, a code must be verified to function as it is intended. This involves, first, successfully compiling and linking and passing vendor/developer-provided tests. While this might confirm a code will run under certain conditions and for certain inputs, it does not guarantee that the code runs *correctly* or that it runs or runs correctly for the intended application. The test criteria might be flawed, or the nature of provided tests may not by applicable, e.g., scale of the calculation—tests are generally smaller problems to enable run quickly—or combination of physics—general purpose codes with a multitude of options cannot test every combination of options—or untested features—actively developed codes with sophisticated features may not provide tests of every feature. Adequate code verification will typically require evidence developed by a user that the code is operating properly, for the intended use. In addition, adequate code verification may require monitoring all calculations performed within the subcontinuum activity—not just pre-existing tests—and confirming the absence of exceptional conditions or warnings. The nature of a failure condition must be specified. A fundamental practice is that all exceptional conditions pertinent to the intended use must be resolved. Not all exceptions are relevant to a given application. Failures of functionality not relevant to a given application might be ignored, but should be identified as a limit of the application. If a failure of verification of a needed function cannot be resolved, ultimately perhaps due to a lack of access to the source code, this might indicate the need to adopt alternate methods to perform the simulations.

Practices associated with code performance verification include:

- **Acquire named, versioned, configuration-managed code.**
  This includes acquisition of the necessary input materials models, atom models, or other managed library of data needed for use of a code. For example, in a DFT code, atomic pseudopotentials specific to the composition of the materials to be studied are a necessary input, or, in an MD code, the interatomic potentials.

- **Successfully compile, link and generate executable on selected platform(s).**
  Ideally, this build process is automated, with reliable report of build success or failure. A substantive record of the build procedure (in the form of makefiles or build scripts) should be maintained, and evidence of success documented.

- **Perform vendor(developer)-provided tests, and resolve failures.**
  Ideally, this test process will be automated, with a reliable report of failures. In

practice, a user must verify functioning of the code(s) through sufficient tests to cover the intended use. Ideally, all failures *should* be resolved, but in practice only failures associated with intended use *must* be resolved. If failures not associated with intended use are unresolved, they must be noted, and further use of the code in the subcontinuum activity must be monitored to ensure that these failure conditions do not occur. Examples of failure modes that can be monitored include: limits of functioning (number of processors, problem size, memory per processors), failure of an (currently) unneeded feature, or failure to achieve convergence in some iterative procedure.

- **Design and execute additional tests to verify code operation for intended use.** Vendor provided tests might confirm that a code runs, and perhaps give the same results that the vendor obtained, but do not prove the code performs the mathematical operations it claims to execute and provide the operation required for the intended use. The user must independently verify the proper functioning of the code. These can take the form of conservation tests, or boundary condition tests, or benchmark calculations. For example, a DFT code solves a set of differential equations in an iterative process to achieve self-consistency to a variational problem. This imposes a stringent set of conditions upon the solutions, which, if satisfied, conclusively demonstrate verification of significant parts of the code. Self-consistency must be achieved; the self-consistent electronic structure solution must have a lower energy that any of the trial functions in the sequence of iterations leading to self-consistency. The forces on the atoms are evaluated in a different set of code as gradients of the energy, and the relaxation of the atomic structure to a ground state following the forces must lead to a structure with lower energy than any of the trial configurations leading to the relaxed structure. These are internal consistency conditions that, in effect require two different sections of code to agree exactly. Electrostatic boundary conditions at slab models are surfaces should be confirmed to isolate artificial images of slabs from each other, this can be done through cell-convergence studies that increase the distance of vacuum between slabs. These kinds of tests must be defined for all M&S activities, and any failure conditions resolved. Conditions or operations that cannot be tested must be listed as assumptions.

- **Specify failure conditions for routine operation of code, and monitor results**, Codes will occasionally fail in run-time, detecting an internal failure condition, or jobs will crash because of platform or system failures. These will be resolved in the normal course of activity, failure signaled by an incomplete calculation. In cases where automated procedures are used to process results, provision must be made to detect these system failures to ensure that interim results of a failed calculation are not propagated into further analyses. Completed calculations are not necessarily successful calculations, and results should be monitored to confirm that no exceptional conditions are encountered. To extend the earlier example, DFT calculations commonly are used to obtained optimal equilibrium atomic structures for a model system. Calculations for the intended use should be monitored to ensure that these relaxation studies do satisfy these success conditions. This will often require manual intervention and expert judgment, and, once again, automated processing needs to be implemented cautiously.

- **Record association of results with code and environment.**
  A practice must be in place to establish provenance for every calculation in a subcontinuum analysis. Record of the code used can be in a version date stamp with output files and log files in the code, or in records developed in the course of M&S activities.

A comprehensive verification of every aspect of the performance of a code is not necessary. For the purposes of code performance verification, only those aspects of code performance related to needs over the range of intended (or anticipated) use must be verified. Moreover, much of the normal operation of the code may be verified through satisfaction of vendor/developer-provided tests, or significant literature that documents specific uses of the code.

3.4.1.3 Physical/numerical model and resource adequacy

Code provenance, and code performance and numerical model correctness are the aspects of code verification that indicate that a given code operates correctly, but is not sufficient to demonstrate that the proposed simulations with the code are adequate for the intended use, as indicated in the "Testing for Adequacy" section of the NEAMS Waste IPSC V&V Plan. A series of questions must be addressed:

- Do the verified features of the code span the anticipated needs of the activity?
- Do the results converge well enough, is the code sufficiently robust?
- Do the verified performance limits of the code (number of atoms, size of meshes, parallel scalability) meet the anticipated needs of the activity?
- Is the available platform(s) adequate to the anticipated needs of the activity (memory/processor, memory/node, node-hours, disk space)?
- Is the total computational resource adequate for the activity, either in throughput or in total cycles over the extent of the activity?

The verification of adequacy with respect to these questions is primarily covered in practices outlined in code performance verification and solution verification, in that the practices there are pitched to verify suitability for the "intended use". These simply highlight the questions that determine decisions made about resource choice and resource allocation for subcontinuum activities, and will involve project management. The verification practices specific to verification of adequacy here are:

- Determine requirements of activity.
- Screen and assess alternatives of codes and methods for intended use.
- Select alternatives based on verified best suitability to requirements and resources.

One example of this practice in action was a recent exercise in the NEAMS Waste IPSC to generate atomic structures for glass surfaces, to characterize chemical interactions with water at the surface [Criscenti2011], ultimately to construct a multi-process model of glass corrosion and assess the limiting mechanisms for glass dissolution. The computational expense of generating glass surface structures from first-principles DFT is prohibitive. Classical MD methods, in principle, are more efficient means to generate randomized structures like glass. However, different MD codes and associated sets of interatomic potentials have different capabilities, such as available potentials for elements, or type of chemistry those potentials types are capable of, that needed to be screened to determine the adequacy to the purpose of generating glass models with the appropriate chemistry, for candidate glass systems with particular atomic composition.

3.4.2 Solution verification

The goal of solution verification is to establish the confidence in the accuracy of the numerical solution generated by the code with respect to a given problem. The problem statement must be free of errors in the problem setup, in the execution of the calculations, and the communication of the results into post-processing and propagating into upscaling into coarser-scale models. The numerical settings, discretizations (meshes), cutoffs, and other flags and options that affect the numerical results must conform to the physical model to be modeled and demonstrated to produce meaningful results with respect to the target output quantity. This typically means that the calculation is cast in an asymptotic form and the settings place the calculation(s) in the asymptotic convergence regime. The numerical errors due to the various cutoff and numerical approximations must be small enough to satisfy the requirements of the intended use, and smaller than the physical errors given by the nature of the physical approximations of the tool. Practices associated with solution verification can be categorized in groups corresponding to (1) input verification, (2) numerical verification, and (3) output verification.

3.4.2.1 Input/output verification

Was the input to the calculations correctly constructed, did the specified job actually correspond to the desired calculation? Did the desired calculation have the correct configuration to represent the desired simulation? Were the boundary conditions or environmental conditions correctly incorporated in the computational model? Was the correct material model or other input invoked, and what is the performance of that input model? Is the output correctly captured? Did the calculation run correctly? Practices for input and output verification include:

- **Verification of the input file (mechanical) by analyst or peer review**
  To confirm that the input is properly constructed to perform the desired calculation.

- **Record the input file and dependencies**
  The purpose is to enable future assessment, or reproduction of the result. A record is sufficient if it enables a deterministic reconstruction of input files necessary to reconstruct the output quantities of a simulation.

- **Document purpose of calculation in input file, or associated records**
  To identify the intended purpose of the calculation, so that the metric for any future assessment is more immediately evident.

- **Verification of appropriate boundary conditions and environment**.
  Many codes have options and capabilities for specifying different boundary conditions, zero-dimensional (in DFT, an isolated molecule in three-dimensional space) vs. three-dimensional (in DFT, a periodic replica), open vs. closed boundary conditions. Are these properly invoked, and are they working properly?

- **Verification of the materials models used and proper implementation**
  Codes are dependent upon inputs of materials models, which themselves need to be verified and used properly. An example might be investigation of defect properties in a material using DFT. The "materials models" in this case would be the choice of functional (constitutive model for electrons) and the choice of atomic pseudopotential. Verification of the model might include computation of equilibrium lattice parameter for the material within the model, and then the use of that lattice parameter in calculation of defect properties, rather than using an experimental lattice parameter (which would potentially corrupt the results by introducing spurious strains). Performance of the materials models and input is as important to the outcome of the calculations, as is the numerical verification of the code. This practice extends to other properties that might be limited by materials model, and overlaps model validation.

- **Verification of correct extraction and recording of the output quantity(ies)**
  This practice involves verifying the process for extracting simulation output quantity(ies) from the calculation. This practice could involve recording the output file(s), or considering the impracticality of storing the voluminous data produced my many codes, a reduced record generated by an electronically driven process (grep, or script, or reduced listing output) that captures the essential elements of the output solution. Records must be kept sufficient to identify and reconstruct the output quantities from the calculated outputs.

- **Verification of correct operation of code (code performance verification)**
  The practice extends the code performance verification practice of confirming proper performance of the code. A record of proper termination, e.g. successful satisfaction of a code termination condition such as a convergence criterion, and evidence that code performance did not experience a verification failure are useful. In a DFT calculation for a structural relaxation to obtain a minimum energy atomic configuration, this might involve documentation that the largest force on any atom is below some specified threshold, and, furthermore, that the sequence of trial configurations has energy minimized to some energy threshold (without a verification failur, e.g., an increase in energy as the structure relaxes following the forces). The calculation might need to satisfy other conditions in order to be meaningful. These criteria should be specified as part of the subcontinuum V&V planning, verified by the user/analyst, and, ideally, saved and recorded within an evidence management system.

3.4.2.2 Numerical adequacy verification

This set of practices requires identifying the code settings or numerical approximations that can affect the numerical solution, and verifying that those settings are adequate to render the numerical results meaningful.

- **Verify material model inputs**
  The quality of a calculation and confidence in the simulations is dependent upon adequate materials models. Is some intrinsic resolution setting adequate? Is the construction of the model adequate in its parameterization? Is the model fully internally consistent? For example, in DFT calculations pseudopotentials (PP) are often used to represent atom potentials. These PP are often distributed with codes as libraries, but frequently users will have opportunity to generate or modify one for their own use. Results of DFT calculations can be profoundly affected by the manner in which these are constructed, with varying degrees of fidelity depending on the specification of a modest number of input parameters. Verification of the specification of the PP—what are the core electrons, what is the atomic valence electron configuration used in constructing the PP, what PP method is used, what are the various cutoff values—would be required practice for a DFT calculation. The functional forms of interatomic potentials and their parameterizations would be an analogous material model input for a classical MD code. The provenance and performance of these material inputs to codes must be identified and verified.

- **Basis, discretization, or representation of solutions**
  The representation of the solutions to M&S must be verified. For example, in quantum chemistry of DFT methods, discretization of solutions for electron orbitals and densities, "basis sets", can be as combinations of plane waves, whose quality is specified via energy cutoffs, or linear combination of atomic orbitals, whose quality is expressed in terms of discrete descriptions such as "valence double zeta". In molecular dynamics, trajectories of positions and velocities of particles are the representation of a solution, but the particles might be individual atoms, or unified clumps of atoms. The quality of these representations should be assessed through resolution-refinement, verified against asymptotic limits or appropriate benchmarks.

- **Computation domain adequacy**
  Verify that the scale of calculations enabled by the code and computational limitations is sufficient for use. This scale may take the form a computational model size (e.g. number of atoms in a DFT supercell) or whether use of 1D or 2D modeling to represent 3D phenomena is sufficient, or whether the length of simulated physical time is adequate to provide meaningful quantitative output.

- **Grids, meshes and associated numerical quadratures, time steps**
  Verify that various real-space, reciprocal space, atomic grid, domain resolution, time step resolution are sufficiently resolved to be sufficiently accurate, fine enough to be within an asymptotic regime with meaningful results. Tests might

take the form of increasing or decreasing the resolutions. In a DFT code, one has an assortment of real-space grids, reciprocal space sampling, and other numerical quadratures that can be manipulated by a user, and these should be verified to be suitable for the intended purpose.

- **Cutoffs**
  Verify that various cutoffs produce results in asymptotic ranges that minimize their affect on the output results. To obtain good scaling properties and optimal computational efficiency, many codes impose cutoffs on particle interaction lengths, integral evaluations, etc. to some threshold. Those cutoffs to which the calculation proves sensitive and must be modified to meet requirements must be identified and verified.

- **Criteria for solution convergence**
  Use of convergence criteria is intrinsic to many codes. Many codes have iterative solutions with thresholds for completion, or simulations for accumulation of statistics for averaged properties (e.g. Monte Carlo methods) expect user-specified criteria for convergence or completion. These choices should be identified and verified. For example, in a DFT calculation, there is a convergence criterion for the self-consistent solution of the electron density, a threshold maximum force that indicates the location of a minimum (or other critical point), or threshold stress for a bulk structure optimization. These thresholds must be tested and verified adequate for the intended purpose [Mattsson2005].

- **Initial conditions in optimization calculations**
  The initial conditions for calculations can determine the final output result of calculations, materially changes the value of an output quantity, In calculations sensitive to input conditions, the intent in selecting the initial conditions with respect to different solutions must be verified. Many subcontinuum scale (and coarser scale) calculations are intended to identify critical points (typically minima) or optimal path on a potential energy surface or for some other objective function. A DFT calculation seeking the ground state energy for an atomic defect in the lattice could very easily be “trapped” in a local minimum because of symmetry constraints, or because the initial starting structure is not in the attractive well of the global ground state. Similarly, a barrier energy to a reaction, such as the breaking of a bond at a glass surface is the lowest energy path from reactant to products is dependent upon specifying and assessing the correct path. Calculations for a ground state structure or lowest energy path presuppose a candidate minimum and are then computed starting from an initial guess configuration that places the solution within that local well. The nature of the search for alternative minima—systematic, random, ad hoc—should be specified.

- **Other parameters that affect calculations**
  Codes have a multitude of setting and parameters that are important to affect the output results. These will be code and domain-specific, and a V&V Plan must identify these and demonstrate a plan for verifying their adequacy. An example from DFT calculation are such parameters for a (artificial) temperature for

electron states and occupation or mixing parameters, which can have a tangible effect on numerical results, and therefore need to verified and documented [Mattsson2005]. In a MD calculations, the choice of thermostat and the means by which is configured (equilibration, temperature ramps, statistics gathering) are settings and parameters important to a calculations that need to be verified.

A comprehensive verification of every input into a calculation is not necessary. The expectation is that generic code performance verification will implicitly verify much of the usual operation of the code with conventional settings. These are not within a problem-specific numerical adequacy verification. The focus of numerical adequacy verification are those parameters and settings that the user *might* have cause to manipulate with specific intent in the usual course of configuring a calculation for its intended use, and which might have consequential effect on the numerical results.

The leading priorities are those settings that for the purpose of the calculations *must* be generally specified by the user in an input and are actively invested in the design of the calculation. Other important priorities are those nominally *optional* settings that a user sets deliberately—or avoids deliberately—to accomplish some goal, either for design of the calculation or in response to a verification failure. The intent, the figure of merit, and verification of the figure of merit must be provided. Other parameters or settings that might play a consequential role, or for which involved conventional use of the code for related simulations or by other users in related simulations, even if not invoked, *should* be noted and verified.

The goal of numerical verification is *not* to prove the fidelity of the algorithms and numerical approximation to analytic purity or machine precision. The minimum threshold is to verify that numerical results are obtained to a *precision* that is meaningful for the purpose of the calculation, that numerical errors or variability do not overwhelm the output quantities and make the calculations valueless. The most stringent useful threshold is determined by the limits of the *accuracy* of the physical approximations expressed by the mathematics. It is futile to refine precision of numerical results significantly finer than the accuracy of the model form. Marginal benefit to reducing an overall uncertainty with further numerical refinement is small: the overall uncertainty of the calculation will be determined by the physical accuracy of the model (which will be the purview of model validation).

3.4.3 Model verification and post-processing

The immediate numerical results of a calculation or series of calculations often does not represent the output quantity(ies) of the activity associated with that calculation. In a process not entirely dissimilar to upscaling or multiscaling, the results are collected, aggregated, filtered, manipulated, visualized, etc: put through a process to obtain the specific output quantities required of a phenomenon. This may involve a manual analysis, simple post-processing companions codes shipped with the code package, or use of sophisticated code packages. In a V&V plan, any such process or model must be identified, its assumptions documented, and the results verified.

The process of manipulating the numerical output of a calculation(s) to obtain an output quantity involves an analysis, often invoking a model. The model, its form, its parameterization, and the process of manipulating the model must be verified. The assumptions implicit in the model should be noted.

Report of experimental observables almost always involves a model and processing. For example, raw data makes for insightful diagrams, but the assumption of discrete processes and the assumption of Arhennius behavior in a model enables conversion of raw experimental data with temperature dependence to be converted into activation energies for those discrete processes. For the purposes of insertion into a NEAMS Waste IPSC data/process flow (for validation), these models and processes of analyzing experimental data are subject to the same requirements of verification and validation as M&S activities, in order to develop quantitative confidence in the output "prediction" of phenomena properties, the quantities that are the end goal of the activities.

Similarly, in a report of computational observables, the output quantity(ies) of M&S activities involves a model and processing raw calculation data into output quantities characterizing the phenomena representing the intended use. An example is constructing statistical or thermodynamic quantities from a trajectory generated during a MD calculation. A more complicated example is adapting a calculation to account for boundary conditions or environmental conditions that are not in the calculation explicitly—this involves some element of extrapolation. Examples in DFT are incorporating pH effects into calculations of energetics of surface chemical processes, or extrapolating results from finite size/time calculations to infinite size/time asymptotic results. The model assumptions, the model form, and the parameters associated in converting results of the code calculations into predictions of phenomena or process behavior should be specified and verified.

The activities associated with this model verification practice, in principle, would be captured in a decomposition of the subcontinuum activity into a system of codes and analyses each as a separate activity. The purpose of calling out model verification as a separate practice is the frequent and customary embedding of these assumptions and models into subcontinuum domain calculation. The distinct activity of analysis subsumed and "hidden" in the context of a given subcontinuum scale activity. This practice requires that a subcontinuum activity identify and call out the embedded analyses and models, make explicit the implicit assumptions, and define the verification of the process in the V&V plan associated with a subcontinuum activity, and subject the crucial post-analyses to the same level of scrutiny as activities associated with M&S code calculations.

3.4.4 Model validation

The goal of model validation is to establish confidence that the predictions of properly verified M&S activities give sufficiently accurate description of reality for the intended use, and provide estimates as to how well that reality is predicted.

The realization of validation is often expressed as the match between experimental results and observations and M&S results, but, as noted above, experiment provides an imperfect view of "reality" for the purposes of comparison. Considerations of verification, validation, and uncertainties apply as much to the experimental view of reality as to the M&S view of reality. The mapping onto models that express our quantitative understanding of reality can be as much a challenge for experiment as it is for modeling and simulations. Hence, a validation exercise is often (usually) more than comparing the calculated M&S result with "measured" experimental result, it is the comparison of a the M&S model of reality with an experimentally originated model of reality.

Validation is not accomplished by the comparison of a single experimental result with a single M&S result (or via a small set of undifferentiating comparisons), but rather is developed as the accumulation of evidence from multiple comparisons over a range of data values that systematically build a quantitative confidence in the predictive power of the capability. A single agreement between a phenomenon characterized via experiment and via M&S (or a narrowly focused set of characterization) could be happy coincidence, perhaps indication that a phenomenon has been qualitatively *described*, but is generally inadequate to demonstrate that the phenomena characteristics are adequately *predicted*, and certainly insufficient to answer the question of how well the target process is predicted, a fundamental goal of validation.

The unknown—frequently unknowable—error in the physical approximation(s) in subcontinuum simulation methods complicate the assessment of validation. Subcontinuum scale simulations are more prone to be limited by errors of model form than they are to errors in numerical form, or, restated, the numerical errors are usually more controllable and assessable than errors in physical approximations. It is the physical approximation that fundamentally limits the *accuracy* with which a process can be simulated, and hence determine how well a model can be validated.

A prominent illustration of this principle within DFT, that validation is mostly a measure of the physical approximations, is in the formulation of DFT itself. The exact electron-electron interactions are not computable for all but the most trivial of problems. DFT replaces these detailed interactions with a functional of the total electron density, an approximate constitutive relation. Many different functionals are available for use in DFT, with differing reputations for accuracy in different situations, and will provide different results for any given chemical system. The DFT error for simulating any given chemical process cannot be predicted. Nonetheless, use of functionals in DFT simulations over a large variety of materials systems and processes has allowed practitioners to develop a collective sense of where a given functional succeeds and where it fails, and, moreover, enables a realistic and often defensible estimate of how badly or well it succeeds for a given class of problems. In molecular dynamics, similar insight into the accuracy achievable with a given interatomic potential is developed within a community through repeated use of candidate potentials. Validation of a physical approximation is intrinsically empirical, and greater confidence in the predictions using a given physical approximation is developed with a larger and more differentiating sampling of comparisons to reality.

Validation of M&S observables is limited by the availability of appropriate experimental and observational data, and, furthermore, the degree of validation is limited by the quality of data. Generally, a M&S capability generally cannot be validated to a finer degree than the degree of uncertainty present in experimentally-derived quantities being compared. A fundamental practice that needs to be developed in any V&V plan is to lay out a plausible approach to obtaining appropriate data, of sufficient quality to meet the validation requirements of the activity.

In general, NEAMS will not generate or commission experimental studies, so these needed data will need to be obtained from other sources. Much experimental and observation data has been accumulated for nuclear waste disposal, and the lore collected there has been used to develop a comprehensive collection of FEPs and PIRTs for the NEAMS Waste IPSC activities. A challenge specific to subcontinuum scale (as opposed to PA or continuum scale) activities is that this lore is mostly restricted to system level quantities at the continuum scale, and generally contain little data usable for validation of subcontinuum scale activities, with the related consequence that the FEPs and PIRTs—expressing requirements—do not cascade downward to the subcontinuum scale. A fundamental practice for the design of subcontinuum activities is the availability of experimental data appropriate to validation of the specific model activities within the line of sight of the activity.

A V&V plan for a subcontinuum activity must include the following practices associated with model validation:

- **Specify validation needs and define protocols**
  Identify phenomena and quantities associated with the M&S activity that must be validated in order to develop a quantitative confidence in the output quantit(ies).

- **Identify source(s) of model-appropriate experimental data**
  Where will the necessary data be obtained? If it is not available and cannot be obtained, can the M&S capability be validated by other means? If not, what uncertainties does this introduce into the analysis, how defensible are the assumptions that must be made in the absence of direct validation?

- **Record data**
  The Data Acquisition practices in the NEAMS Waste IPSC V&V Plan [NEAMSWaste2011] outline practices associated with acquiring, assessing, and recording data for NEAMS activities.

- **Reduce numerical uncertainties to less than physical uncertainties**
  Validation is only meaningful if the numerical errors in the subcontinuum are demonstrated to be less than the errors given by the physical approximations, or at least that both the numerical and physical uncertainties are shown to be less than the uncertainties in the experimental results for the physical property.

- **Perform validation comparisons**
  Typically, a single point validation is not reliable. A small set of comparisons has

greater credibility. A systematic comparison over a larger data set is desired, to allow a statistical assessment of validation. If different physical approximations are possible and plausible within a given subcontinuum simulation method, systematic comparisons between the results of different methods gives greater credibility to the validation, and also can contribute to a judgment concerning the level of error in the physical approximations. The goal is not just if the M&S activity is giving an acceptable result, but to provide defensible quantitative measures of how right or wrong the results are for the intended purpose.

- **Resolve all validation failures**
  Validation failures encountered in executing the validation plan can be due to failures in the model, inadequacy in the physical approximations, failure in the numerics, or failure in the code, and thereby cast doubt on the entire M&S capability, and must be resolved. It may be that the experimental data point is wrong, it may be that the comparison between experiment and M&S is poorly constructed or misinterpreted, or it may be that the validation point is less directly applicable to intended purpose of the activity. These latter conditions might indicate that the validation protocol needs to be reformulated rather than the M&S capability disqualified. The record of the validation failure and its resolution should be recorded, to denote the problem with the data, or to delineate the limits of applicability of the M&S capability, or to appropriately note a weakness in the validation of the M&S, with the associated justifications and explanations.

In summary, the goal of model validation is to create a body of evidence that makes a convincing case that the models and codes produce results that are a sufficiently accurate description of reality to meet the requirements of the intended use, and to provide defensible estimates as to how well that reality is predicted. The practices outlined in a V&V plan should ensure that all crucial aspects of the model are tested, that those tests are meaningful, and thereby establish quantitative confidence in the M&S capability.

3.5 Uncertainty quantification and sensitivity analysis

Confidence in results of subcontinuum scale activities requires explicitly understanding, identifying, controlling, modeling, and quantifying the error and uncertainties in the system of methods, codes, models, data, and calculations used in the analyses to generate the output quantities of the intended use, and demonstrating that those errors and uncertainties can be made or are sufficiently small to meet the requirements of the intended use. Uncertainty quantification and sensitivity analysis are mechanisms used to assess and establish this level of confidence. Practices to enable quantitative assessment of the confidence in the output quantity(ies) of the intended use must be a part of a V&V plan for a subcontinuum scale activity.

Uncertainty quantification (UQ) is the systematic process of identifying the sources of errors and uncertainties in subcontinuum activities, and aggregating these into a meaningful *quantitative* measure of the total level of confidence of the output quantity(ies), meaningful in the sense that defensible decisions can be made concerning the adequacy of the result.

Sensitivity analysis (SA) is the systematic process of identifying the degree to which output quantities are affected by changes to specification of the inputs to a calculation (typically, but not necessarily only, parameters). An effective SA identifies the most important inputs to a simulation and informs optimal prioritization of efforts.

Conventionally, uncertainties are categorized into stochastic (aleatory) uncertainty, an intrinsic random variability in a process or phenomenon, or state-of-knowledge (epistemic) uncertainty, stemming from an incomplete or approximate understanding of the physical phenomenon. At the subcontinuum scale, these concepts map most constructively and usefully into numerical uncertainties—controllable uncertainties determined through specification of input parameters affecting the *precision* of a calculation—and physical uncertainties—irreducible errors in the physical approximations, a model form error that limits the *accuracy* of a simulation of a physical process. A viable V&V plan for a subcontinuum scale activity must address both numerical and physical uncertainties in a meaningful and balanced fashion.

3.5.1 Physical uncertainties

At the subcontinuum scale, the physics/chemistry models are frequently "first principles" or deterministic, in the sense that there are few or no free *empirical* parameters to manipulate. The leading, most important uncertainty is the fidelity of the discrete choice of model form, e.g. a specific flavor of density functional in a DFT calculations or the selection of a particular form of interatomic potential in an MD simulation, or in the choice to use quantum chemistry rather than a classical interatomic potential approach.

Different methods and codes have different strengths and weaknesses for different applications, commonly understood by their practitioners. For the purpose of assessing the total uncertainties in a subcontinuum activity, realistic, useful, and defensible estimates of the errors of the physical approximation applied to the intend use must be generated. The uncertainty in the physical approximations used in a model determine a useful metric for uncertainties in the activity, and will inform decisions concerning V&V and UQ procedures for the activity, and determine fundamental choices concerning the activity itself. For instance, if the limit of physical uncertainty is shown to be inadequate to the requirements of the intended use, the approach to obtain the intended output quantity must be reconsidered, perhaps, e.g., requiring a quantum chemistry approach to replace a classical approach to quantify a chemical process.

Because the model form error is irreducible, the assessment of uncertainty (error) in the output quantity due to the physical approximation is inherently empirical. This necessarily involves expert judgment and domain expertise.

A comparison between of a single experimental observable and a computational observable (or a small set of undifferentiating comparisons) provides some plausibility that a physical model can *describe* a physical process, but has limited value for *predicting* how well a physical model describes that process, *i.e.*, estimating the errors in that physical model.

Quantitative confidence in the prediction of an output quantity, extending to defensible estimate the physical uncertainties, is developed through the detailed and systematic accumulation of validation evidence. Greater and more varied comparisons against data are needed to establish broader confidence in the M&S capability. Greater statistics in the comparisons allows more defensible estimates of the errors inherent to the physical approximation.

The availability of validation data is an important component to making that assessment. A plan for subcontinuum scale activities must include the identification, acquisition, and recording of validation data sufficient to make a quantitative assessment of the physical uncertainties.

Other, less quantitative, probes of the confidence in the results of a calculation are to vary the choice of physical approximation among different plausible choices. The sensitivity of the results to changing the flavor of functional (or pseudopotential, etc.) used in DFT calculations gives a qualitative sense of theoretical error. The greater the manipulation of the physical approximations, and the less the sensitivity of the result to those physical approximations, generally the greater confidence one can place in those results. However, this is a qualitative judgment, not a quantitative one, and while perhaps desirable in certain applications, the detailed comparisons to data are to be accorded greater credibility than detailed comparisons between different M&S results.

The first priority of any subcontinuum activity is to demonstrate suitability of purpose for the intended use. The first step in that demonstration is the evidence that the errors in the physical approximation are, or can be made, small enough to satisfy requirements. This places a fundamental lower limit on the uncertainty possible in an activity.

3.5.2 Numerical uncertainties

Sources of numerical errors and uncertainties are identified in the normal practices of code performance verification and numerical adequacy verification. Numerical uncertainties arise due to compromises made in the computational model, such as discrete integration grids or reciprocal space samplings to evaluate three-dimensional integrals in a DFT code, or the finite length or time scales used in MD simulations to collect statistical information for the evaluation of thermodynamic quantities, or they can arise in specification of the material model inputs to the calculation. A variety of parameters that potentially define a given calculations, as described in 3.4.2.2, can contribute to numerical uncertainty in the ultimate desired output quantity.

The total numerical uncertainty due to sensitivity of the output to these input numerical parameters must be adequately assessed.

The minimal threshold is to verify that all parameters are sufficient to achieve a meaningful calculation for the intended purpose. Establishing this minimum threshold is the purpose of verification practices described above. Meshes must be sufficiently refined to be in an asymptotic regime, cutoffs set to reasonable values, convergence criteria sufficient to give at least coarsely converged results.. This assessment is typically informal, dependent upon the informed expert judgment of an analyst, and asserts a qualitative assessment of the uncertainties. The standard is to establish a *plausible* suitability for purpose, useful for scoping studies and initial sensitivity analyses targeted to evaluating uncertainties due to specific input parameters.

A comprehensive UQ involving assessment of uncertainties due to each input into a calculation is not necessary, an adequate assessment addresses just those inputs that contribute significantly to the overall UQ for the intended use. The focus for UQ are those settings and parameters that a user has cause to manipulate with specific intent in the course of configuring a calculation for its intended use, that are demonstrated or are suspected to have a consequential effect on the numerical results, that cannot or will not be refined to have a negligible effect when applied to compute the quantity(ies) of interest.

The most stringent useful threshold is to demonstrate that the numerical uncertainties associated with the calculations are much smaller than the physical uncertainties. It is pointless to refine numerical uncertainties to significantly finer than the limit of the accuracy of the physical approximations. It is good practice to develop defensible estimates of physical uncertainties before investing significant resources into evaluation of numerical uncertainties, to enable optimal decisions for resource allocation devoted to UQ of numerical uncertainties.

The target threshold is defined by the uncertainty requirements of the intended use. The exact requirements may not be well defined before or during significant subcontinuum activities. Thegoal of subcontinuum activities is to develop and refine a continuum model of coarser scale phenomena, and the sensitivities and associated requirements are only characterized after that output model has been assessed. The exercise of the Waste IPSC system on a specified application will propagate requirements down from the system scale, ultimately providing refined requirements for the global system into uncertainties within the local line-of-sight.

A useful practical threshold is to target refinement of contributions to the numerical uncertainty that is balanced between different contributions, focusing priorities on the largest contributors to numerical uncertainties.

The initial step is to assess the sensitivities of the output results to the parameters and settings outlined in numerical adequacy verification practices (3.4.2.2). This begins with an informal assessment driven by expert judgment, particularly in exploratory and scoping phases of subcontinuum work, progressing to systematic assessments as further knowledge is developed and ithe level of rigor advances.

Many of the sources of uncertainty can be made very small with appropriate choices of setting and parameters, with respect to other sources of numerical uncertainty or with respect to physical uncertainties, for settings that will be used in simulations. These contributions to uncertainty should be verified to be below some bound, through convergence studies, benchmark comparisons, or other systematic investigation of sensitivity.

Those sources of uncertainty that cannot be reduced below some insignificant bound, must be assessed and defensible evaluations of uncertainties performed.

Meshes and grids can be refined and demonstrated to exhibit convergence to an asymptotic limit, and the variability in the sequence of resolution-refinement studies used to evaluate or bound an error for the mesh or grid to be employed in the production simulations. Cutoffs, criteria for solution convergence, and other tunable parameters can be similarly assessed, varying the input parameter or setting and showing convergence to an asymptotic limit, or sensitivities below some bound, and using the observed variability, at the settings planned to be used in the production simulations, in the desired output quantity as an evaluation of a numerical uncertainty. For DFT calculations, this would include uncertainties due to k-point sampling, or electronic temperature, or basis set specification in the construction of meaningful calculations of selected materials properties

Where refinement studies are impractical, benchmark comparisons, to analytic solutions or appropriately designed test cases that are numerically converged, can be used to estimate uncertainties. The degree to which numerical approximations are seen to violate known conservation or internal consistency rules are also relevant to uncertainties, and must be included in an assessment of overall uncertainties.

Uncertainties due to initial condition specification (and settings with similarly discrete numerical consequences) are more difficult to assess quantitatively. Levels of confidence in these aspects of the M&S capability must be developed through the accumulation of evidence that a search space has been adequately sampled to find a global minimum, or some other useful bound on the magnitude of the uncertainty can be developed, or via comparisons to appropriate benchmark problems.

The source of numerical uncertainties must be identified, and then verified to be insignificant, or the uncertainties explicitly evaluated, or estimated from an accumulation of evidence, in any total uncertainty quantification associated with a subcontinuum activity.

3.5.3 Sensitivity analysis

Uncertainty quantification is used to measure the level of confidence attained in an output. Sensitivity analysis (SA), conversely, is used to measure the dependence of an output on the specification of an input, measuring the level of confidence needed in an input. The conceptual purpose of SA is to identify weaknesses in a model and quantify the degree of weakness, for the purpose of either efficiently prioritizing efforts within the activity to reduce overall simulation uncertainty or, alternatively, propagating well-defined requirements into a subordinate scale that defines materials models or other needed input.

In addition to being necessary to guide activities within the subcontinuum model, both credible and defensible UQ and SA are vital outputs for subcontinuum activities to be incorporated into system model development and upscaling. For upscaling, UQ—degree of confidence of the output quantities—is needed to quantify the degree to which requirements are satisfied in an upscaling process. Conversely, SA—the degree of confidence needed in the input quantity—is needed to define requirements for upscaling from a lower scale. Subcontinuum scale activities respond to SA from continuum scale simulations, where identification of quantitatively crucial phenomena drives requirements and priorities for subcontinuum scale activities.

Sensitivity analyses and uncertainty quantification are often simultaneous, where an analysis of sensitivity, *e.g.* how sensitive is the result to a mesh refinement, is also the assessment of the associated uncertainty, *e.g.* the variability in the output as a result of mesh refinement (evaluated at the mesh refinement to be used in the application).

3.5.4 Progression of V&V and UQ activities

Sensitivity analyses begin with expert judgment based on domain expertise. Experienced practitioners typically will have a native sense of the parameters important to a simulation, and a semi-quantitative understanding of the minimal requirements in those parameters to obtain a meaningful result. This level of SA is typically associated with exploratory, scoping studies. Among the goals of such a scoping study could be a preliminary assessment: if the simulation approach is adequate to describe a phenomenon, if the phenomena is likely to be important, to get a preliminary assessment of the physical uncertainty present in the description of the phenomenon with the approach.

In the normal course of code performance and numerical adequacy verification practices, the sources of uncertainties in the input to a calculation will be identified. The sensitivities to all parameters are verified to be small enough to produce a meaningful result. For the most sensitive parameters, those predicted to have a significant effect on the output quantities and deemed to have the largest effect on total uncertainties, the SA will expand into a quantitative evaluation of the uncertainty (*e.g.*, mesh refinement) and demonstrate that this satisfies, or can be made to satisfy, a physically meaningful threshold. This level of SA and associated UQ are typically associated with simple description of phenomena.

To establish a predictive capability, quantitative assessment of uncertainties must be extended to all parameters and settings that contribute significantly to uncertainties, as identified via systematic SA. Systematic validation is required to obtain defensible quantitative estimates of physical uncertainties, and defensible aggregations of numerical and physical uncertainties combined into an overall uncertainty quantification. These elements are required of activities supporting model construction intended to make predictions, predictions that can be assessed to have certain level of confidence. This level of SA and UQ would be associated with construction and qualification of candidate upscaled models to be used in coarser scale simulations.

With a predictive, assessable capability established, the performance of the new model can be assessed at the coarser scale for its intended purpose at that scale, and sensitivities to the subcontinuum-resolved components of the model can be resolved. Assessed inadequacies in the model at the continuum scale are propagated downwards to the subcontinuum as new requirements, imposing a specified level of confidence on the components of the model. Response to these requirements, in refinement of the model validation and uncertainties and demonstration that the uncertainties satisfy the required threshold, requires detailed identification and control of all sources of uncertainties.

As subcontinuum activities advance along this natural progression, from feasibility and plausibility, to predictive accuracy for specific intended use, the requirements for V&V and UQ also advance, from informal judgment to detailed justification and documentation. The specific requirements associated with advancement through these progression will be domain specific—for example, DFT codes and meso-scale phase field applications will have very different paths to establish quantitative confidence associated with a given level of rigor—and also application-specific. But every subcontinuum activity will advance through this progression, and a V&V plan for that activity must define a series of requirements reflecting that progression.

3.6 Evidence management

Management of appropriately retrievable V&V evidence and documentation of UQ and SA is crucial to establish quantitative confidence in an M&S capability toward its intended use. The purpose of V&V is to inform meaningful decisions concerning the activity, and the nature of the audience and purpose of the decisions, the levels of rigor required, determines the requirements concerning the amount of data capture and the enduring quality of data capture.

For exploratory scoping studies, the purpose is to explore suitability of purpose and provide a preliminary assessment of the feasibility and plausibility of a given M&S approach to a proposed materials property characterization. This is typically informal with a very qualitative purpose: is the approach worth pursuing further, is the expected physical uncertainty likely to be adequate for the intended purpose, is it computationally tractable? This process is guided by an analyst, is dominated by expert judgment, and informs a short term decision, with little risk, of whether to continue the approach. The level of data capture required is minimal: what methods were considered and the basis for the qualitative judgment. The information needs to endure only sufficiently to make that decision, with sufficient record in the program to motivate the decision.

For subcontinuum characterization of individual phenomena, the importance in the overall Waste IPSC PIRT for a system performance assessment is usually unknown, and part of the purpose of the subcontinuum investigation is to determine *if* the phenomenon plays an role, if the property satisfies some threshold behavior, or perhaps it is one of many competing candidate phenomena for a coarse scale behavior. The quantitative characterization must be convincing and the evidence caprure sufficient to make the characterization credible. This the domain of typical scientific research, and a journal publication or project report presenting the result and summarizing the research might constitute sufficiently documented evidence. The documentation must provide enough details to be assessable. In principle, the full specifications necessary to make the results assessable would be presented in the publication (or associated supplementary evidence). At the least, all evidence associated with verification and validation should be recorded at the project level, and be available to be presented on demand, in a local evidence management system or repository. Ideally, within NEAMS-chartered activities, this evidence would be captured in a formal IPSC-wide EVidence Information Management (EVIM) System, but the results of any activities that are queried through literature or other sources should have sufficient documentation for traceability and reproducibility, from which it would be possible to assess the quality of the results (any inadequacy flagged in the recording system).

To qualify the characterized phenomena for inclusion in an integrated upscaled model, a candidate coarse scale (constitutive) model, all sources of uncertainties must be assessed (at least with expert judgment), and evidence of verification and validation and UQ and SA assessments captured and recorded. Qualification evidence for the model must be available to a larger audience, now involving continuum-scale efforts, The evidence must be more enduring, as the proposed model will be used for assessment of the performance of the constitutive model in continuum M&S, and potentially become part of the system performance assessment. The wider audience and the need to document a more global assessment mandate more data be captured and stored as the quantitative support for the mode. The evidence must endure and be accessible by an audience that must be presumed not to have access to the original authors. Ideally, this would be in an IPSC-wide EVIM, but all the data must be captured in some enduring repository accessible to the entire IPSC team, so that the data could be migrated or registered in an EVIM by the team in the event that the constitutive model passes muster and qualifies for use in continuum simulations.

Model certification activities, where the model has been demonstrated to be important in an overall system performance assessment, and which has propagated requirements downwards into subcontinuum-resolved components of a constitutive model, entail meeting specified thresholds of accuracy and uncertainty on the subcontinuum results. The results must be comprehensively assessed, verified, validated, with UQ and SA, to the extent possible, the evidence accumulated and recorded, and placed in an enduring, queriable repository. With requirements from the performance assessment level now explicitly propagated into the subcontinuum scale, the phenomenon characterization will have entered a system-wide PIRT. The activity inherits all the requirements associated with the demanded, including being subject to independent review, This requires enduring, independently retrievable evidence storage and management.

A V&V plan for a subcontinuum activity must describe a plan for data capture and appropriate evidence management and reporting commensurate to the level of rigor of the activity, and, furthermore, provide for migration of the associated V&V evidence into enhanced data management environments as the results of the activity advance through this progression. The details of what constitutes adequate evidence to establish confidence in a capability at a given level of rigor will differ with the domain and the specific application, but simple guidelines considering basic questions—considering the audience, the purpose and nature of the decision being informed, the need for independently accessible documentation—are common questions that must inform the requisite standards for evidence management.

4 V&V and UQ Assessment

The typical life cycle of a subcontinuum investigations directed toward the development of improved constitutive models based on first-principles mechanistic process can be described as follows.

(1) Assess suitability of purpose – approach plausible, feasible, to meet requirements? This scoping stage involves a preliminary assessment of physical uncertainty, and determination that the goal is computationally tractable.

(2) Characterize selected process – construct quantitative prediction of property? Predict quantities with defensible accuracy, and assessable uncertainties. Assess physical uncertainty through validation, verify numerical uncertainties sufficient small to have physically meaningful results, assess largest uncertainties.

(3) Install process into upscaled model – sufficient to support qualified model? The results should be fully verified and validated, with defensible UQ assessments of output quantities, to a standard sufficient for a continuum scale assessment of the model to make discriminating assessments of the subcontinuum-resolved components of the constitutive model.

(4) Support certified model – refine model to specified UQ requirements. All significant sources of uncertainty assessed, and refined to satisfy thresholds specified by system-level requirements.

The degree of assessment of verification and validation, and of quantitative uncertainties for a subcontinuum activity must be tailored to the level of rigor demanded by the purpose of each level in this life cycle. Mapping this life cycle onto a Predictive Capability Maturity Model provides a useful classification scheme for identifying a level of rigor associated with a given activity, and also the standards for V&V and UQ and evidence management.

A useful mapping onto a PCMM for subcontinuum scale activities occurs within the line-of-sight of the activity, requirements for V&V and UQ are pitched to the immediate purpose of the activity. Requirements for evidence management are determined by the anticipated audience for the V&V evidence. This section summarizes the mapping onto a level of rigor, and then the practice-resolved requirements for a given level of rigor.

4.1 Defining level of rigor

The level of rigor associated with the subcontinuum activity is determined by an assessment of the purpose of the activity, within the line-of-sight of the activity, and the intended or anticipated audience, as summarized in Table 4-1.

Table 4-1. Determining the level of rigor

Activity	Purpose	Output	Audience	Level of Rigor
Scoping, exploratory	Determine suitability for purpose of approach	Preliminary physical uncertainty, computational tractability	Analyst (self)	0
Phenomenon characterization	Quantitative prediction of processes, assess role in constitutive model	Meaningful, defensible, validated quantitative predictions, assessed physical uncertainties, verified assessable numerical uncertainties.	Peers and colleagues	1
Incorporating process into constitutive model development	Insertion of process into upscaled model, for qualification of candidate improved constitutive model.	Quantitative predictions, fully validated, assessed physical and numerical uncertainties.	IPSC Team	2
Constitutive model refinement and certification support.	Systematic refinement of model accuracy and uncertainties to specified required thresholds.	Fully validated quantitative predictions, with fully assessed physical and numerical uncertainties, refined to meet specified requirements.	Independent review	3

4.2 Evidence management

At all levels of rigor, a minimum requirement for evidence is documentation of sufficient detail concerning the activity to enable independent assessment of the confidence to be placed in the results. The collection and generation of V&V evidence and UQ assessment must be sufficient to satisfy the level of rigor demanded of the activity. The storage and retrieval of evidence associated with that level of rigor depends on the audience that needs to be convinced, and the extent to which that they need to be convinced. Every bit of data need not be stored—code can generate gigabytes and even terabytes of data during the operation—but the immediate results of an assessment must be accessible, and the recorded data stored as evidence must be sufficient render the results reproducible, traceable, and assessable. Table 4-2 summarize the nature of evidence management required at each level of rigor.

Table 4-2. Evidence management versus Level of Rigor

Level	Standard	Documentation	Evidence retrieval	Storage type
0	Plausible	Informal	From analyst	Informal
1	Convincing: independently assessable and reviewed by peers	Peer-reviewed publication, supplementary material, project reports for additional evidence	Results from documentation, supporting evidence from analyst on demand	Enduring, analyst access
2	Quantitative: assessed by analyst, reviewed by Team.	Above, plus formal Team review	Results and supporting evidence from Team storage, detailed supporting calculation records from analyst.	Enduring, Team access
3	Predictive: assessed by analyst, assessed by Team, reviewed by Independent Party	Above, plus subject to formal independent review	Comprehensive supporting evidence and details from EVIM.	Permanent, independently accessible

Enduring, for the purposes of this classification means a repository that is accessible, queriable, and meaningful to its intended audience, that has sufficient redundancy to prevent loss of data after an unfortunate event. Ideally, all appropriate data from subcontinuum activities for an intended use could be captured in a fully functional EVIM as described in the Waste IPSC V&V Plan. More practically, an analyst might keep detailed electronic records for a project in personal directories on a local disks of a compute server, whose data is backed up to remote servers, or a Team keeps a detailed record of activities on a wiki server accessible to the entire Team, also systematically backed up to remote disks—these would satisfy the minimum requirements of storage. In the absence of a fully functional EVIM, the requisite evidence must be collected, classified, and recorded in an enduring storage type appropriate to the intended use.

4.3 Code verification

Necessary and acceptable code verification practices will vary from code to code, whether the code in question is a commonly used commercial code, or a Waste IPSC-developed capability, or a nameless analyst-developed code with little documentation. A commercial code may not make source code available to users, a private analyst code may be in a state of continuous development with little or no documentation. Satisfying the requirements of code verification for a given level of rigor will entail different practices for these different kinds of codes. The goal is a provable level of reproducibility and assessability of a capability for a specified level of rigor, and documented satisfaction of verification with that capability to a specified level of rigor.

This might entail adding version control, configuration management, and a distribution/recording mechanism for private analyst codes to enable provenance and reproducibility, or more comprehensive verification tests and benchmarks to more openly available equivalents for a closed source commercial code; either path must make provision for developing a requisite level of confidence in the code for the intended use. Table 4-3 summarizes plausible assessment criteria for code verification appropriate for different levels of rigor, adapted for subcontinuum scale activities.

Table 4-3. Code Verification versus Level of Rigor

Level	Code provenance	Code performance	Verification test plan	Review and Documentation
0	Code acquired, supporting libraries and environment defined.	Code builds and runs	Informal or none.	Little or none, based on analyst judgment
1	Code identification and provenance recorded, supporting libraries and environment recorded	Passes vendor tests, and analyst-designed verification tests, resolving all application-specific failures	Informal, analyst judgment, based on vendor tests and analyst defined tests.	Peer and colleague review, of publication, supplementary material, and project reports.
2	Code provenance and platform-dependence recorded, code and environment, or verified equivalent, independently assessable	All code performance issues identified and documented, verified to be satisfied.	Includes a test suite vendor tests and analyst-designed test for application-specific issues.	Verification test plan, reviewed by peers and by Team.
3	Code provenance and platform-dependence recorded, code and environment, or verified equivalent, independently reproducible	All code performance issues identified, tested, and documented to be verified, including run-time and post-analysis error detection.	Full test suite, including verification of run-time and post-analysis verification failures.	Formal verification test plan and recording of verification results, subject to Independent Review.

In the absence of detailed control of the code development process, i.e., lack of a strong formal regimen of software quality engineering—the norm for the dynamic and diverse community in subcontinuum material and chemical modeling—careful code verification practices targeted to the intended use are particularly important to specify and document.

4.4 Solution verification

Necessary and acceptable solution verification practices will differ even more strongly from domain to domain, and application to application. The scale of criteria for solution verification presented in Table 4-4 are one acceptable mapping of solution verification consistent with levels of rigor, adapted to a generic subcontinuum scale activity.

Table 4-4. Solution Verification versus Levels of Rigor

Level	I/O Verification	Numerical model adequacy	Materials model (input) adequacy	Criteria
0	Casual	Physical meaningful; by expert judgment.	Expert judgment	Expert judgment: Estimated numerical uncertainties comparable to estimated physical uncertainties
1	By Analyst	Semi-quantitative; identification of all sensitivities, investigation of dominant sensitivities.	Relative performance of different models investigated.	Augmented by explicit evaluation of dominant numerical uncertainties, demonstrated less than detailed validation of physical uncertainties
2	By Peers	Quantitative, with systematic evaluation of sensitivity to dominant numerical inputs and investigation sensitivity of solutions to all numerical inputs.	Sensitivities identified, documented, and controlled to extend practical.	Verified assessment of numerical adequacy uncertainties to a quoted level of confidence
3	Reproducible, subject to independent review	Predictive. Systematic refinement of all numerical inputs found to affect solutions.	Systematic SA performed, documented, and propagated as requirements for lower scale	Full assessment with quantitative uncertainties below specified threshold.

As in the other V&V activities, the criteria.for solution verification must be tailored to the level of rigor, the classification for a subcontinuum domain and application guided by the purpose and audience of the activities given at a given level of rigor.

4.5 Model and analysis verification

The immediate output of the code is frequently must be filtered through a model or post-analysis to obtain the quantitative output of the intended use of the activity. Ideally, these models and analyses would be called out as separate activities subject to independent

V&V and UQ assessments in a chain of distinct subcontinuum activities, but often are so embedded in a computational application of a given domain (example: analysis of raw experimental data to distill quantities associated with conceptual processes) that it is not practical or worthwhile to make a distinction. The output quantities of the calculation can depend on the formulation of model or analysis, and therefore activities associated with model and analysis deserve targeted attention and subject to the same standards of rigor in a V&V and UQ assessment as the aspects dominated by computation. These models and analyses typically embed important assumptions. There assumptions must be explicitly identified and convincingly demonstrated to be valid to establish confidence in the overall M&S capability. Table 4-5 illustrates model verification practices.

Table 4-5. Model analysis verification versus Levels of Rigor

Level	Model and Analysis Verification Activity	Goal
0	* Identify model post-analysis assumptions * Specify analysis flow	* Plausibility
1	* Specify model forms and validate * Verify analysis flow * Develop test plan – requirements for acceptance	* Credibility; Passes qualitative tests
2	* Verify model assumptions, convergence to correct limits, verify correct analysis * Assessed UQ and SA against test plan	* Quantitative; Passes quantitative tests
3	* Model/analysis flow and model parameters/process verified and validated * UQ and SA assessed to specified threshold	* Predictive; Subject to Independent Review

The criteria are necessarily vague, because of the varied nature of model analysis. Criteria will be specific to how prominent a role a given post-analysis plays in a particular application. Any V&V and UQ plan must identify the model, their assumptions, a test plan for verifying the assumptions and correct analysis flow, and specify criteria consistent with the need to establish confidence in the analysis consistent with the level of rigor of the combined subcontinuum activity. As a general principle for establishing more specific criteria for model post-analysis verification, the best reference point is to consider the post-analysis as an independent subcontinuum activity, with its own full V&V and UQ requirements to establish quantitative confidence to a prescribed level of rigor, and then mapping that mental exercise onto this condensed set of criteria.

4.6 Model validation

For subcontinuum activities, faithfulness of the physical approximation to reality is typically the most significant source of error. Careful and detailed validation is crucial to determine the magnitude of these errors—the physical uncertainties. Close integration and coordination with a significant experimental effort is crucial to success of an M&S activity, and availability of data will drive key decisions in M&S approaches.

In a validation-bound activity, validation criteria associated with each level of rigor must be selected to provide the greatest useful guidance for the next stage in the progression. The scoping stage must identify usable data for validation and set a defensible bound for physical uncertainties, in advance of any process characterization. Without adequate data, the validation required to advance through levels of rigor is impossible; without a credible sense of physical uncertainties, the numerical uncertainties to adequately describe a materials process cannot be described; and if the approach has physical uncertainties larger than what is required of the total uncertainties, another approach must be chosen entirely. Table 4-6 illustrates escalating validation criteria that step-wise inform prioritization and resource allocation advancing through the progression of levels of rigor.

Table 4-6. Model Validation versus Levels of Rigor

Level	Model Validation Activity	Goal
0	* Identify available experimental data * Comparison between plausibly converged numerical result and experiment (or validated benchmark)	* Preliminary bound on physical uncertainty
1	* Acquire appropriate experimental data * Multiple comparisons of credibly converged numerical results to wide sampling of experimental data * Identify validation gaps (existence of data for certain processes, or experimental uncertainties unknown?) * Develop validation plan, cognizant of state of experimental data	* Semi-quantitative bound on physical uncertainties
2	* Acquire differentiating experimental data with uncertainties for validation * Systematic comparison of assessed-converged numerical results to differentiating set of validation data. * Identify validation limits (where are weaknesses?)	* Quantitative assessment of physical uncertainties
3	* Acquire differentiating validation data with well-characterized experimental uncertainties * Quantitative comparisons of predictive accuracy for all output quantities	* Full predictive capability to specified requirements

4.7 Uncertainty quantification and sensitivity analysis

Table 4-7 summarizes a progression of criteria consistent with a given level of rigor, that mostly restates the requirements embedded in the V&V criteria describes above.

Table 4-7. UQ and SA versus Levels of Rigor

Level	UQ/SA Activity	Responsible party	Review	Decision informed
0	Approximate physical uncertainties. Numerical uncertainties estimated to be in physically meaningful regime	Analyst, using expert judgment	Little, none.	Is approach physically plausible and computationally feasible?
1	Evaluate physical uncertainties. Numerical UQ for (assumed) dominant sensitivities bounded to be in meaningful regime.	Analyst, using expert judgment augmented by numerical tests and validations selected by judgment-ranked dominant sources of uncertainties.	Peers and colleagues.	Has the approach achieved credible accuracy for process, and what would be needed for that accuracy and its uncertainties to be fully assessed?
2	Physical uncertainties fully assessed, all significant sources of numerical uncertainties assessed or verified to be bounded to less than physical uncertainty. SA for input materials models identified and estimated.	Analyst, using detailed comprehensive analysis, with systematic investigation of UQ to all significant parameters, and detailed validation. Team, through review and assessment of model in continuum M&S.	Team.	Has the capability matured sufficiently to be used in an engineering assessment of an upscaled model incorporating the modeled process?
3	Uncertainties comprehensively analyzed, with no significant assumptions. SA for input materials models assessed and propagated as requirements to lower scale.	Analyst. Team, through review and assessment of model in continuum M&S, with subcontinuum-resolved SA.	Indep. Review	Can the modeling of the process be refined sufficiently to meet performance assessment requirement with desired confidence?

5 Path Forward

Analyses performed with subcontinuum-scale activities to support NEAMS Waste IPSC assessment will require establihsing confidence in the subcontinuum M&S capabilities commensurate to the risks associated with decisions that the analyses will support, consistent with the level of rigor determined through application of the requirements line-of-sight criteria. The principles illustrated in this section should be incorporated into V&V and UQ planning for every subcontinuum activity, to ensure that the M&S capabilities meet these requirements. The progression of practices with increasing levels of rigor parallel the practical life-cycle of subcontinuum investigations targeted to an application, and merely express the standard of due diligence that are expected for results to be able to be sufficiently trusted for a particular intended use. The purpose of V&V and UQ planning for a subcontunuum activity is to chart a reasonable and achievable path of continuous improvement to satisfy the incrementally rising standards demanded of results as they are used in making and supporting decisions of increasing consequence. In each subcontinuum domain, this path will traverse different issues of greatest concern, but the need to establish and document measurable criteria tailored to different levels of rigor is a common requirement. The goal is articulating an achievable path through a domain-specific landscape of verification and validation challenges.

This V&V and UQ document is not a V&V plan. It articulates a strategy and expectations for developing V&V plans for subcontinuum activities in support of a Waste IPSC performance assessment.. Like the Waste IPSC V&V Plan it descends from, it outlines a framework for building a V&V plans, but focused on subcontinuum M&S and explicitly mindful of the need to inform upscaling as the immediate intended purpose of the activities.

Implementation of V&V and UQ practices into individual subcontinuum activities is a prerequisite for developing a systematic, quantitative, and *assessed* approach to upscaling, bridging between levels in the system hierarchy. The input to an upscaling process must be assessed, satisfying a known level of confidence. Similarly, the subcontinuum activity must identify and assess its sensitivities to material model inputs, in order to be able to define meaningful requirements on the lower scale processes and models. The development of achievable, meaningful V&V and UQ plans for subcontinum activities and implementation of the appropriate practices will provide a basis from which upscaling is constructed, and is the foundation from which to develop approaches to assess and propagate uncertainties through the unsolved and poorly characterized filter of upscaling.

Implementation of these V&V principles should begin upon initiation of any subcontinuum activities engaged in support of the NEAMS Waste IPSC. In the early development of the Waste IPSC, most subcontinuum efforts will be early in the life-cycle, in scoping and prototyping. The V&V and UQ requirements associated with these levels of rigor are only little more onerous than documenting the due diligence and suitability of purpose that should be the minimum acceptable standard for most scientific activities, adding the perspective of the intended use of the output quantities to a deliberate, systematic scrutiny of the aspects of the activity that affect the output results. The transition from informal to more stringent requirements as the life-cycle advances to higher levels of rigor, and the need to inform decisions of greater consequence, involve incremental escalation of V&V requirements.

References

Criscenti, L.J., et al. (2011). *Progress Report on Subcontinuum Scale Glass Corrosion Challenge Problem*. SAND2011-8250. Albuquerque, NM: Sandia National Laboratories.

Knowledgebase of Interatomic Models, (KIM). (2011). http://openkim.org/

Kuksin A.Y., I.V. Morozov, G.E. Norman, V.V. Stegailov, and I.A. Valuev. (2005). "Standards for molecular dynamics modeling and simulation of relaxation." *Molecular Simulation* 31, nos. 14–15, 15-30: 1005-1017.

Mattsson, A.E., P.A. Schultz, M.P. Desjarlais, T.R. Mattsson, and K. Leung. (2005). "Designing Meaningful density functional theory simulations in materials—a primer." *Modeling and Simulation in Materials Science and Engineering* 13: R1-R31.

NEAMS Waste Forms Team:. (2009). *Waste Forms and Systems Integrated Performance and Safety Codes System Design Specification*. SAND2009-3969. Albuquerque, NM: Sandia National Laboratories.

NEAMS Waste Forms Team: Freeze, G., J. G. Arguello, R, Howard, J, McNeish, P. A. Schultz, and Y. Wang. (2010). *Challenge Problems and Milestones for NEAMS Waste IPSC*. SAND2010-7175. Albuquerque, NM: Sandia National Laboratories.

NEAMS Waste Forms Team: H.C. Edwards, J.G. Arguello, Jr., R.A. Bartlett, J.F. Bouchard, G. Freeze, R. Howard, P. Knupp, M.T. McCormack, P.A. Schultz, A. Urbina, and Y. Wang. (2011). *NEAMS Waste IPSC Verification and Validation Plan (Version 1)*. SAND2011-0084. Albuquerque, NM: Sandia National Laboratories.

Oberkampf, W. L., M. Pilch, and T. G. Trucano. (2007). *Predictive Capability Maturity Model for Computational Modeling and Simulation*. SAND2007-5948. Albuquerque, NM: Sandia National Laboratories.

Distribution

(to be distributed electronically)

1	MS0899	RIM-Reports Management	9532 (electronic copy)

Sandia National Laboratories